Award Number: W81XWH-16-1-0460

TITLE: Applied Cognitive Models of Behavior and Errors Patterns

PRINCIPAL INVESTIGATOR: Dr. Robert Wray

CONTRACTING ORGANIZATION: Soar Technology, Inc.
Ann Arbor, MI 48105

REPORT DATE: September 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2017	Annual	15 Aug 2016 - 14 Aug 2017

4. TITLE AND SUBTITLE

Applied Cognitive Models of Behavior and Errors Patterns

5a. CONTRACT NUMBER

5b. GRANT NUMBER

W81XWH-16-1-0460

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Robert E. Wray, PhD.

wray@soartech.com

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Soar Technology, Inc. (SoarTech)
3600 Green Court, Suite 600
Ann Arbor, MI 48105-2588

8. PERFORMING ORGANIZATION REPORT NUMBER

10315

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command

Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY ACRONYM(S)

USAMRAA

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This effort's over-arching goal is to study, to model, and to apply predictive markers (indicative behaviors) during medical training, focusing on application of the markers when the learner makes observable decisions (pivotal opportunities). We are investigating the activity patterns that learners exhibit while interacting within learning scenarios. Activity patterns include the timing of decisions, and observations of mouse movements, button clicks, and dwell times. Learning scenarios are situated in training for Emergency Medical Technicians, focusing on the cognitive, perceptual and affective knowledge and skill that is necessary for "sizing up" an accident or incident scene on first arrival. The effort has two specific aims: 1) Develop training scenarios that present pivotal opportunities and elicit indicative patterns of behavior from learners; 2) Develop computational constraint-based models of indicative patterns. This report summarizes progress and accomplishment toward both aims.

15. SUBJECT TERMS

computer-based learning, adaptive learning, behavioral patterns, emergency medical technician.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	None	57	USAMRMC
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER *(Include Area Code)*

TABLE OF CONTENTS

1. INTRODUCTION

This effort's over-arching goal is to study, to model, and to apply predictive markers (*indicative behaviors)* during training, focusing on application of the markers when the learner makes observable decisions (*pivotal opportunities*). We are investigating the activity patterns that learners exhibit while interacting within learning scenarios. Activity patterns include the timing of decisions, and observations of mouse movements, button clicks, and dwell patterns. Learning scenarios are situated in training for Emergency Medical Technicians and focus in particular on the cognitive, perceptual and affective knowledge and skill that is necessary for "sizing up" an accident or incident scene on first arrival. The effort has two specific aims: 1) Develop training scenarios that present pivotal opportunities and elicit indicative patterns of behavior from learners; 2) Develop computational models of indicative patterns. This report summarizes progress and accomplishment toward both aims.

2. KEYWORDS

Computer-based learning, adaptive learning, behavioral patterns, emergency medical technician (EMT), mouse-tracking, behavioral indicators

3. OVERALL PROJECT SUMMARY:

The statement of work for the effort is summarized in Table 1, including a short description of each major task. Note that Tasks 1 and 2 are focused on specific aim 1 (present pivotal opportunities and elicit indicative patterns) and Task 3 is focused on specific aim 2 (develop computational models of the patterns). In the following, we discuss Objectives, Results, Progress and Accomplishments for each task in the Statement of Work.

Table 1. Project Statement of Work

Task 1. Scenario Development
This task is to develop and to validate training content and scenarios. Scenarios are implemented within the Adaptive Perceptual and Cognitive Training System (APACTS). Training scenarios are designed to include supportive, constructive guidance and feedback to present when the learner takes any given action—both for acceptable responses and for erroneous ones. Scenarios are focused on scene size-up for Emergency Medical Technicians. These scenarios involve healthcare content appropriate for an entry-level learner to become familiar with, with a variety of situations portrayed across the entire set of scenarios.
Task 2. Study Design and Data Collection
This task primary focus is to design a study to test the effectiveness of scenarios in identifying behavioral and error patterns in the learning environment and to then conduct the study, collecting and analyzing the resulting data. As an initial step in study design, this task includes an analytic study designed to estimate parameters important for the eventual study design such as the required accuracy of behavioral markers to support effective adaptation for learning based on indicative patterns.
Task 3. Process Modeling
This task is to create models of participant behaviors across the scenarios developed in Task 1.

The models compare acceptable behaviors (such as the correct answer to a direct question) and the indicative patterns that led to a chosen answer (such as the mouse movements and dwell times associated with the choice). The models are also developed to be integrated estimates of proficiency and checks on learning (such as explicit questions). At each pivotal opportunity, where a participant is to make a decision in the scenario, we will extend APACTS to record the participant's actions along with the time, and form an assessment against one or more learning objectives. The resulting history of estimates over performance in the scenario can provide insights into the specific progress of learning.

Task 1: Scenario Development

Objectives and Results
1. Identify sources of training materials.
    - **This objective is met**. We are focusing on the emergency medical technician (EMT) domain, which offers a standardized curriculum on which we can create training scenarios.
2. Develop instructional design for the scenarios.
    - **This objective is met**. We have developed both a complete instructional design and a basic instructional template for each training scenario.
3. Assess and validate the instructional design.
    - **This objective is met**. The standardized, national curriculum has been previously validated and our scenarios hew closely in content with the standard curriculum. We also are engaging subject matter experts in the EMT domain to review specific content presentations, focusing especially on images.
4. Implement the instructional design in APACTS.
    - ***This objective is underway***. We have implemented a few full scenarios to enable testing and expect to complete development of all scenarios (covering the entire instructional design) in October 2017.
5. Encode domain meta-data (learning objectives, expected error types, etc.) in APACTS scenarios
    - ***This objective is underway***. We have extended the APACTS learning environment to support the requirements for responding to behavioral patterns and encoded the learning objectives from the standard curriculum into the APACTS scenarios.

Progress and Accomplishments with Discussion

After search and evaluation of potential options for content, we decided to focus on Emergency Medical Technician (EMT) training and, in particular, one unit within that training. EMT training has several advantages for the effort: 1) curriculum requirements are standardized (*1*), which essentially places some bounds on the role of instructional design within content design; 2) many organizations offer EMT courses and there are many resources on the web about EMT training, which has alleviated some of content-generation constraints and the need for specialized expertise (i.e., in comparison to combat medics) for creation and validation; and 3) EMT programs (including the subset we have chosen) require development of cognitive, perceptual, and psychomotor skill. In the study, we will be focusing on the first two of these, but having

more than one type of skill that needs to be developed should help demonstrate the value of behavioral markers for differentiating learning needs.

We chose to focus on the "Scene Size-up" component within the EMT course. The recommended time for this lesson is 1 hour. Within this lesson there are cognitive, affective, and psychomotor learning goals and the goals include not gaining knowledge but being able to demonstrate and apply that knowledge during the course of the lesson. The relatively short duration of the lesson with a relatively wide variety of learning objectives and types of objectives, makes it a reasonable choice for testing the development of markers, because adaptive choices can potentially focus on choosing alternatives among these categories rather than fine-grained distinctions within a few learning objectives.

The instructional design for the study includes the following units:

- Introduction (What is scene size up?)
- Key Concepts (Introduce terms such as mechanism of injury (MOI))
- Identifying Hazards (general introduction)
- Assessing the complexity of the scene (Can you handle this situation?)
- Vehicle Injuries (general intro)
- Mechanisms of Injury: Front-end collision
- Mechanisms of Injury: Side-impact collision
- Mechanism of Injury: Rear-end collision

For each unit, we have developed an overall template, the structure of which is summarized in Figure 1. Each unit includes some number of introductory "frames" (comparable to a briefing slide) that introduces the topic, terms, and provides examples and explanations. The learner is then presented with a series of vignettes that require a decision/choices. These are the pivotal opportunities in the instructional presentation. What the learner views next is dependent on the choice the learner makes. There are generally five distinct choices:
- Move on to the next item (which could more another pivotal opportunity or new content)
- Reconsider your answer / repeat
- Remediate current topic: Feedback is provided that is focused on the current topic and relatively fine-grained distinctions about the topic.
- Remediated contrasting learning objectives: Feedback is provided that discusses differences between the current topic and/or learning objective (e.g., evaluating potential mechanisms of injury between side-impact and rear-end collisions)
- Remediate concepts: Feedback is provided that focuses on high-level conceptual distinctions, such as the difference between a mechanism of injury (the physical forces that can result in patterns of injury) and the injury itself.

For this effort, these choices are not hard-wired to specific learner responses. Instead, the system uses the computational models of behavioral patterns (discussed further under Task 3) to evaluate which content option is most apt for the current situation. The decision context thus includes the learner's decision/response, the current estimates of skill for the learning objectives relevant to the decision, and the behavioral markers.

Figure 1. The basic structure of APACTS EMT Scenarios in support of the study.
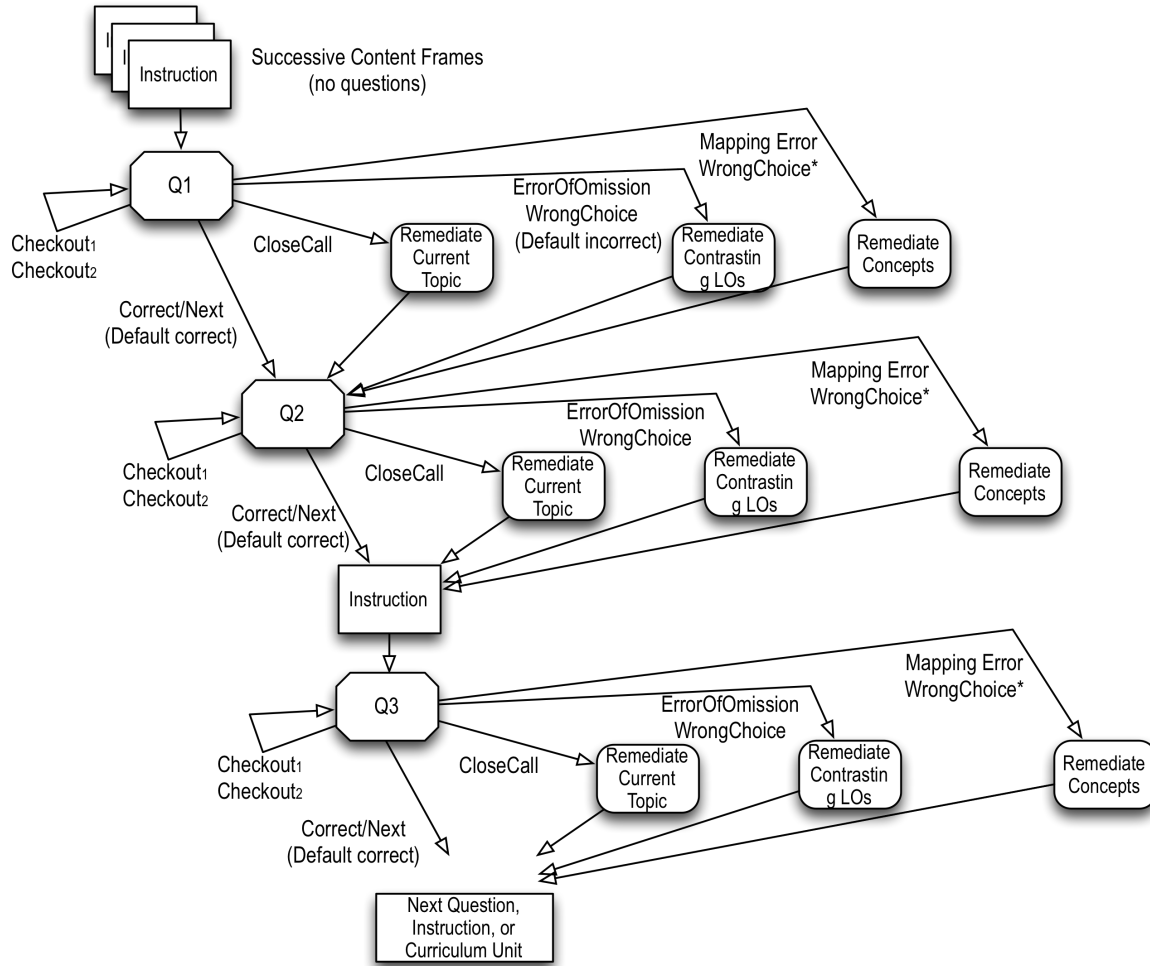
Examples of implemented scenarios are included in Appendix B, the IRB protocol for the primary study.

Task 2: Study Design and Data Collection

Objectives and Results

1. Design and conduct an analytic (verification) study to inform the design of a human-subjects (validation) study.
    o **This objective is met**. The verification study is summarized in (*2*), which is attached as Appendix A. This analysis enabled us to estimate learning impacts across a large space of learning design alternatives. The results of this analysis lead to us to understand that the study required a larger number of content options for each pivotal opportunity and that the study would require a larger number of subjects (about 100) than the original, notional plan (about 50 subjects). The
2. Design a human-subjects study with the goal of investigating the impact of behavioral markers in an adaptive learning environment.

- o **This objective is met**. The protocol for the human-subjects validation study is included in this report as Appendix B.

3. Prepare formal documentation for the study, submit to Institutional Review Board, and obtain approvals from IRB and Army HRPO to conduct the study.
    - o ***This objective is partially met***. The study protocol documented in Appendix B received IRB approval on 21 Jul 2017. The protocol has been submitted to HRPO and is currently under their review.

4. Conduct the study (including subject recruitment, data collection, etc.).
    - o ***This objective is not yet met***. Assuming HRPO reviews are complete, we plan to conduct the primary data collection for the study in Oct-Dec. We have requested a contract modification to enable primary data collection at the University of Alabama but the approved IRB allows collection both at Soar Technology's Florida office and the University of Alabama.

5. Perform data analysis on collected data and summarize overall results and recommendations.
    - o ***Work toward this objective has not yet begun***. We plan to conduct summary data analysis Jan and Feb of 2018.

Progress and Accomplishments with Discussion

The goal of the verification-study design was to establish reasonable bounds on potential learning benefits for indicators in an adaptive training context. The study builds on prior work establishing the use of verification methodologies for the preliminary evaluation of adaptive training systems (*3, 4*).

The study employed a simulated students paradigm (*5-9*) to assess theoretical benefits of more targeted assessment via indicative patterns. A secondary goal of the verification study was to identify an appropriate region(s) along a learning curve for human studies. For example, it may be useful to focus more on intermediate or advanced learners to see a large difference in outcomes than novice learners. These kinds of issues reflect why waiting to design the human subjects study until after the verification study is completed is preferable. The primary results of the study are:

- *Behavioral markers must be highly accurate* to facilitate observable impacts on learning given basic constraints on the study design. The outcomes led us to focus optimizing mouse-tracking before investigating other sources of behavioral markers, as mouse-tracking has been shown to be fairly reliable in many realistic usage contexts (*10*).
- A relatively *large number of alternatives are needed at each pivotal opportunity* to effect observable changes in learning outcomes. The content design takes this factor into account in two distinct ways:
    - (1) We increased the number of content alternatives available at each pivotal opportunity. This change requires more investment in content, but the study showed that having just a few choices at each opportunity was not sufficient for discrimination across the number of pivotal opportunities a learner could complete in 60-90m of learning experience.
    - (2) We designed each pivotal opportunity so that the learner faces choices that correspond to a small number of learning objectives (2 or 3) rather than any learning objective in the curriculum. This approach imposes more constraint on

content development, but ensures that the resulting feedback is targeted to the learner's misconceptions when incorrect or suboptimal choices are made.

- Behavioral markers will have *greater impact and discrimination for novice learners*. Given study constraints, the impacts of behavioral markers will be more much evident (discriminable from the resulting data) if the learner's are not already knowledgable of the domain. This result led us to focus on a more general target population for the study (college students) than a population already familiar medical procedures like medical or nursing students.

A more complete summary of the verification study is included in Appendix A.

Based on the verification study, we designed the human subjects study and documented a protocol for that study. The research compares the results of learning between an adaptive medical learning unit to a unit presented in a non-adaptive (fixed) sequence. As above, the curriculum units focus on "Scene Size Up," a required curriculum component used in Emergency Medical Technician (EMT) training (*1*). These units (both adaptive and non-adaptive) will be presented to university subject population(s) in order to assess the utility of markers to improve adaptive learning in emergency medical environments.

The following variables of interest will be implemented and observed in the study:

- **Instructional approach**: The overall instructional approach of the learning environment. For this study, there are distinct instructional approaches:
  - o Non-adaptive/traditional: An instructional unit that is presented in a fixed sequence to all learners.
  - o Adaptive based on performance (only): An instructional unit in which specific content presentations are constructed/chosen based on learner performance and subsequent estimates of learner knowledge and skill.
  - o Adaptive based on performance and markers: An instructional unit that is dynamically constructed/chosen based on a combination of direct learner observation (as above) and behavior markers.
- **Markers**: Patterns of observed behavior that are hypothesized to have a role in improving a learner model.
- **Knowledge gain**: A measure of the post-test performance of subjects, relative to pre-test performance.

This study is implemented as a between-subjects design, with "instructional approach" being the independent variable of interest. Instructional approach will be manipulated at three levels (as discussed above): non-adaptive, adaptive based on performance (only) and adaptive based on performance and markers.

The primary dependent variable is "knowledge gain", as measured by difference scores between pre- and post-tests given to participants. Additionally, behavioral markers derived from dynamic tracking of mouse movements, will be used to predict learner needs and adapt the learning environment. The combination of these variables will enable the study to address the primary hypotheses, as well as quantify the utility of the chosen adaptive learning models for improving learning in medical environments. The complete study protocol is included as Appendix B.

Task 3: Process Modeling

Objectives and Results

1. Assess modeling options and develop a framework of indicators.
    - o **This objective is met**. We evaluated options and identified mouse tracking as the behavioral indicator of highest priority given study constraints.
2. Define an algorithmic approach for assigning meaning to behavior indicators in the context of the learning environment and interactions among learning objectives.
    - o **This objective is met**. Building from general frameworks for characterizing learning and misconceptions (e.g., *Mind Bugs*) and previous work reifying learning concepts in a practical software implementation, we created a method for assigning meaning/interpretation to patterns of mouse movements and mouse behaviors.
3. Develop models for mouse tracking (primary modeling option).
    - o **This objective is met**. Drawing from the results from the previous two objectives, we have implemented, tested, and verified computational models that perform the interpretation of mouse tracking, recognizing learner patterns and assigning them an interpretation in the context of the current learning situation.
4. Integrate the models in the APACTS learning environment.
    - o **This objective is met**. The models developed under the previous objective have been integrated within the APACTS software for use in APACTS learning environments. This integration included software testing and verification of software functionality of the models within units of learning content.
5. Refine and extend models.
    - o ***Work toward this objective has not yet begun***. We await HRPO review to begin learner assessment. Actual assessment of learners will enable us to identify additional needs and limitations of the models, and to then extend and/or refine of the models based on initial observations and results.

Progress and Accomplishments with Discussion

We evaluated two existing approaches to behavior and error classification: Van Lehn's learner-behavior classification scheme (*11*) and Rasmussen's Skills, Rules and Knowledge (*12*). After evaluation of each of these methods and reference to them in the design of the verification study, we determined to use Van Lehn's *Mind Bugs* taxonomy for classification of errors. This taxonomy is more comprehensive than SRK and while it is also more descriptive than SRK (i.e., rather than generative), we did not identify any major stumbling blocks in encoding recognition rules from the taxonomy in the error recognition system. We have recently extended the framework to include the Knowledge-Learning-Instruction (KLI) (*13*) and the Interactive, Constructive, Active, and Passive (ICAP) (*14*) frameworks. These frameworks take a more current and comprehensive view of learners and learning environments and have facilitated making more fine-grained distinctions in assessment and task contexts for modeling learner behaviors and errors.

For encoding recognizers or "markers" in the learning environment, we have developed models that build on a prior constraint-based behavior modeling system (*15*) to encode non-symbolic behavior patterns. We are focusing primarily on mouse movements and mousing behavior generally as an indicator of both cognitive and affective state. Patterns of mouse movements have reasonable correlation with a learner's affective state (*16*) and multiple studies suggest that learner mouse movements can be effective in identifying learner cognitive state (*17-19*).

Figure 2 summarizes the mouse tracking algorithms, which perform the first step in the recognition process. The learner has been asked to "annotate" the image in the APACTS frame, identifying any objects in the image that is a "hazard" as defined in the EMT curriculum. Positional information is captured, along with the velocity and acceleration of the mouse movement and mouse clicks (represented in the diagrams by the vertical, dashed lines). The velocity and acceleration graphs include examples of both raw (blue) and filtered (green) data. The filters help reduce some of the noise due to inadvertent mouse movements and mouse jitter.

The positions of key objects in the scene are labeled as meta-data (part of Task 1; illustrated in Figure 3), enabling the mouse-tracking algorithm to relate mouse actions to learner activity. For example, in the first and second mouse click events (2nd and 3rd vertical lines in the figures), these areas are associated with the bystanders/potential patients in front of the cars. Although the behaviors appear quite different (compare the two velocity spikes), these are readily classified as comparable outcomes in the learning environment via the use of the labeled areas in the content illustrations.
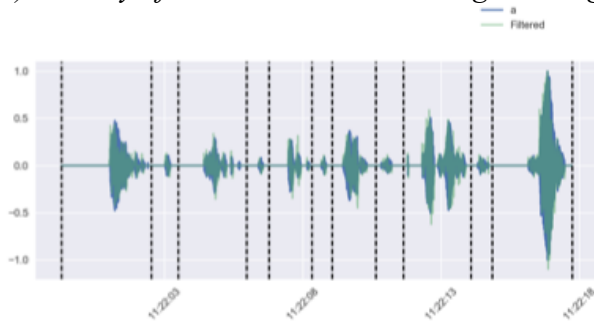

(a) tracking learner mouse movements

(b) (x,y) position of movement
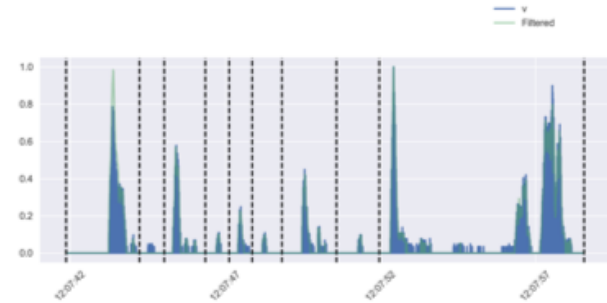
(c) velocity of mouse movement during tracking

(d) acceleration of mouse movement

Figure 2. Basic steps in tracking mouse movement.

By application of Fitt's Law and the filtered data, the tracking algorithms can be used to estimate the confidence of an individual decision. For example, in the latter part of the scenario, the mouse tracks to a few locations but the user does not make a mouse click. By comparison of velocities and accelerations of these different movement patterns, the algorithm attempts to assess the confidence of the learner's decision.

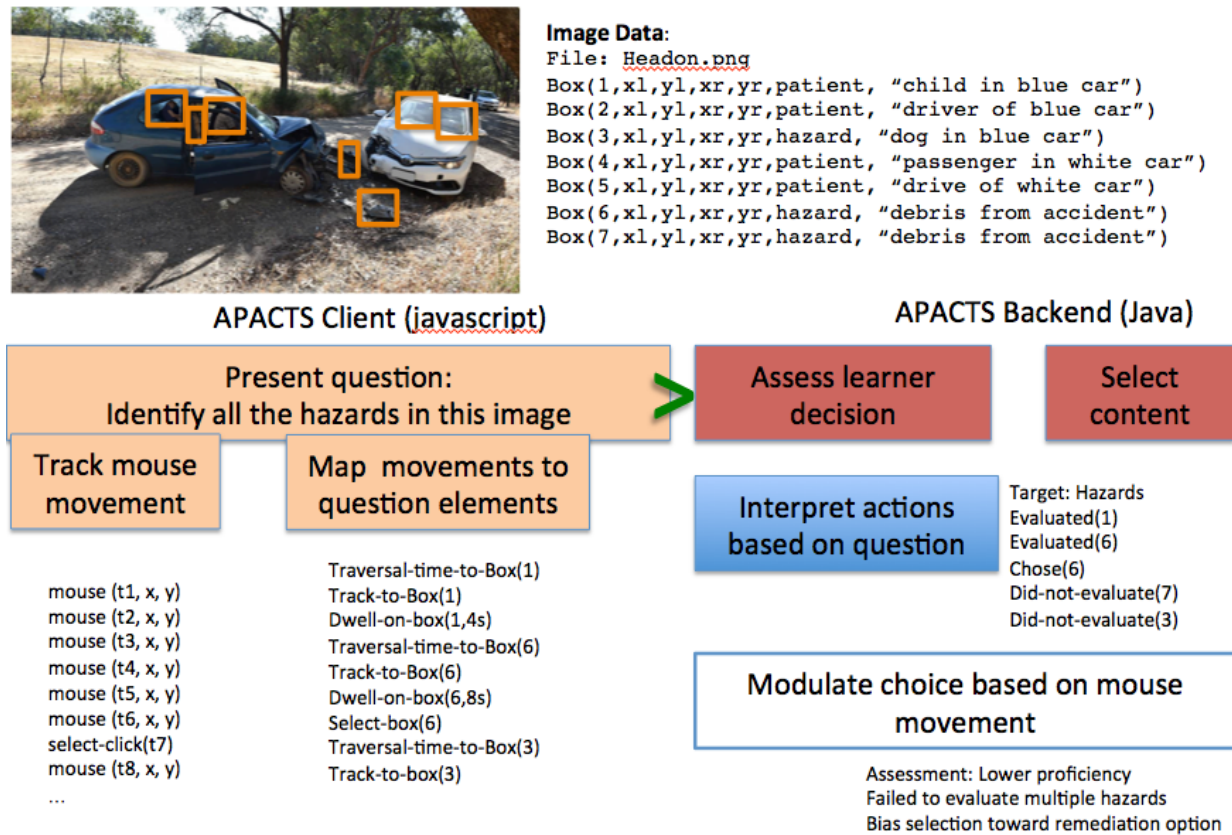*Figure 3. Translating Mouse Movement into Learner Assessment.*

Figure 3 summarizes the information flow that results in these assessments. Following the low-level tracking illustrated in Figure 2 (summarized in the "track mouse movement" component in the figure above), the capture movements are mapped to task interpretations, such as moving to a labeled object ("track to box"), dwelling on a box, and a normalized traversal time. The mouse tracking feeds the primary model (blue component), which focuses on the interpretation and evaluation of the learner's choices. In this example, the model is indicating which of the labeled areas were evaluated by the learner, which of those boxes the learner actually chose, and which boxes the learner did not appear to evaluate based on mouse movements.

These evaluations then feed to the content selection algorithm in APACTS, which determines what content the learner sees next. In the situation shown, the learner's proficiency estimate for relevant learning objectives is low and the mouse tracking lets the system understand that the

learner did not even appear to evaluate hazards in the image. The lack of evaluation results in a bias toward one of the remediation options.

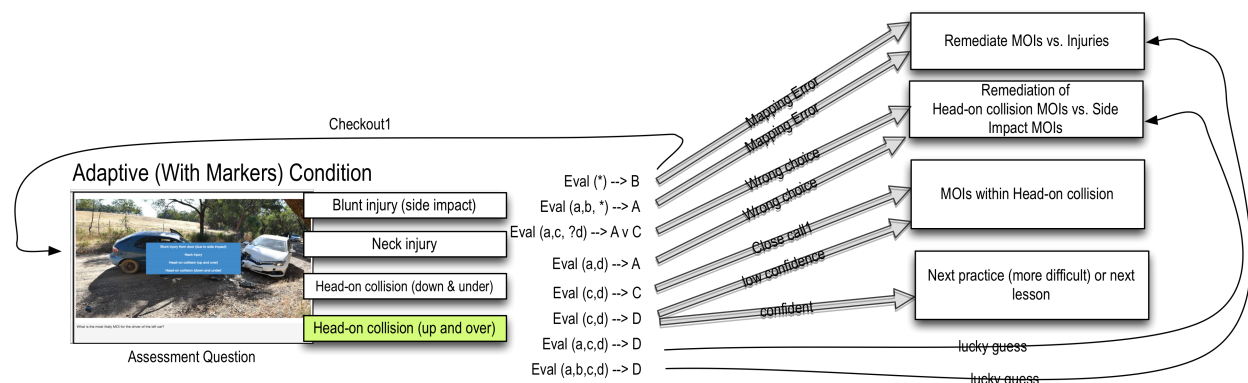*Figure 4. Markers enable alternative tailoring choices for the same answer.*

Figure 4 illustrates an example of the way the model impacts the final content selection decision by the APACTS system. In this multiple-choice question example, the learner is asked to classify the mechanism of injury (MOI). The evaluations of the mouse movements can lead to different responses for the same question. For example, if the learner spends a lot time evaluating all of the options (including item (b), which is a different category of response than the others), the system will choose to remediate MOIs vs. injuries even though the learner's eventual response was the correct choice. The examples in the figure highlight the overall potential value of the markers and models; they provide additional context for interpreting learner activity and tailoring the presentation of content to the learner.

4. KEY RESEARCH ACCOMPLISHMENTS

- Completed verification study, which provided critical insights for designing a human subjects study (Task 1)
- Completed study design (and associated documentation) (Task 2)
- Researched and developed computational models that interpret behavioral patterns and translate those patterns into more fine-grained learner assessments than just the observation of the learner decision provides (Task 3).

5. CONCLUSION

Personalized learning, in which a learning environment adapts to the abilities, needs, and preferences of individual learners, has been identified as a "Grand Challenge" for 21st century research and engineering (*20*). The benefits of adaptive learning environments include more efficient learning (*21*), improved attention and motivation (*22*), the development of less rigid and more flexible decision making (*23*), and improved transfer of learning to settings in which learned knowledge is used and applied (*24-26*).

Improved and personalized learning has particular application for more pervasive and less costly medical training, which often is delivered primarily by human instructors in classes

with modest student-to-teacher ratios. Human instruction and mentoring is very valuable and desirable, but adaptive personalization methods offer an opportunity to deliver good, effective introductory and basic training, thus potentially enabling a single human instructor to train many more students by better preparing them for coaching and instruction from experts.

Adaptation to a learner usually requires a model of the learner that is frequently updated as a learner progresses through a curriculum. Creating a complete and accurate learner model is difficult, however. Markers are designed to improve learner modeling. The model of the learner is frequently updated as a learner progresses through a curriculum (*27*). The targeting of adaptive techniques, such as scaffolding (*28*) and competency matching (*29, 30*) depends on the accuracy (and, to some degree, precision) of the learner model. When the model better reflects the learner's actual knowledge, skills, and attitudes at any point during the learning, the targeting of the adaptive method to the learner generally improves (*29*). Creating a complete and accurate learner model is difficult, however. In addition to estimating learner capability from formal and informal assessment within the environment (*31-34*), researchers have explored many behavioral, physiological, and even neurological indicators or "markers" that can provide additional context for estimating a learner's cognitive state and improving the dynamic assessment of the learner.

Behavioral sensors (posture, eye trackers), physiological sensors (Galvanic skin response), and neurological sensors (EEG) have all been used to assess and track learner arousal/attention in learning environments (*35*). These sensors provide details information but at the cost of introducing uncommon and costly new hardware requirements for the learning environment. However, there is significant and growing scientific evidence that the temporal patterns of mouse movements during selection tasks can provide reliable insight into the cognitive state of subjects (*17, 18*). Mouse-based markers may be noisier (less diagnostically precise) than neuro-cognitive markers associated with specialized sensors but they are omnipresent on standard computer workstations where actual learning environments are deployed. Thus, this study focuses on evaluating the impact of the behavioral markers on the adaptive learning system to improve learning outcomes, taking into account the noise and uncertainty of measure inherent in unspecialized sources.

Our focus commonplace hardware to make behavioral observations, such as a computer mouse, distinguishes this effort from work that uses more specialized sensors to recognize indicative patterns. The study we will be executing over the remainder of the effort will provide insights into the potential benefits (and limitations) of using behavioral patterns derived from everyday and pervasive hardware to improve learning outcomes for medical training. We expect these results to provide evidence of the value of capturing and encoding models of these patterns, and thus providing a foundation for on-going and new learning applications that use models of behavioral patterns to improve learner assessment and targeted of learning content based on those improved assessments.

6. PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS:
    a.   Manuscripts submitted for publication during the period covered by this report resulting from this project:

Wray, R. E., & Stowers, K. (2017). Interactions between Learner Assessment and Content Requirements: A Verification Approach. *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017) and the Affiliated Conferences*, AHFE 2017, Los Angeles.

b. List presentations made during the last year (international, national, local societies, military meetings, etc.).

The peer-reviewed conference publication was presented at the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017) and the Affiliated Conferences in Jul 2017.

7. INVENTIONS, PATENTS AND LICENSES
Nothing to report.

8. REPORTABLE OUTCOMES

The Adaptive Perceptual and Cognitive Training System (APACTS) tool being used on the effort is being used by other projects and groups within Soar Technology for learning sciences research and the development of adaptive training applications. The computational process models (described in Task 3) have been integrated with APACTS are expected to be used in future applications of this software to training applications.

9. OTHER ACHIEVEMENTS
Nothing to report.

10. REFERENCES:

1. United States Department of Transportation, National Highway Traffic Safety Administration. (1996).
2. R. E. Wray, K. Stowers, in *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017) and the Affiliated Conferences, AHFE 2017*. (Springer, Los Angeles, 2017).
3. R. E. Wray, B. Bachelor, R. M. Jones, C. Newton, in *Lecture Notes in Computer Science: Proceedings of the Human Computer Interaction International (HCII) Conference*. (Springer-Verlag, Los Angeles, 2015).
4. R. E. Wray, A. Woods, J. Haley, J. T. Folsom-Kovarik, in *Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics (AHFE 2016) and the Affiliated Conferences*. (Springer, Orlando, 2016).
5. B. MacLaren, K. R. Koedinger, in *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* S. A. Cerri, G. Gouarderes, F. Paraguacu, Eds. (Springer-Verlag, Berlin, 2002).
6. C. J. MacLellan, E. Harpstead, R. Patel, K. R. Koedinger, in *Proceedings of the 9th International Conference on Educational Data Mining-EDM'16*. (2016).
7. K. Van Lehn, in *Advanced Research on Computers in Education,* R. Lewis, S. Otsuki, Eds. (Elsevier, Amsterdam, 1991).
8. N. Matsuda *et al.*, Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology* **105**, 1152 (2013).
9. C. J. MacLellan, K. R. Koedinger, N. Matsuda, in *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*. (Honolulu, 2014), pp. 551-560.
10. M. W. Scerbo *et al.*, "The efficacy of psychophysiological measures for implementing adaptive technology" (National Aeronautics and Space Administration, 2001).
11. K. Van Lehn, *Mind Bugs*. (MIT Press, Cambridge, MA, 1990).
12. J. Rasmussen, Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics (SMC)* **13**, 257 (1983).
13. K. R. Koedinger, A. T. Corbett, C. Perfetti, The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science* **36**, 757 (2012).
14. M. T. H. Chi, R. Wylie, The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist* **49**, 219 (2014).
15. A. Woods, B. Stensrud, R. E. Wray, J. Haley, R. M. Jones, in *Proceedings of the Florida Artificial Intelligence Research Society (FLAIRS) Conference*. (AAAI Press, Hollywood, FL, 2015).
16. S. Salmeron-Majadas, O. C. Santos, J. G. Boticario, in *18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*. (Poland, 2014).
17. E. Hehman, R. M. Stolier, J. B. Freeman, Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations* **18**, 384 (2015).
18. B. Quétard *et al.*, Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision* **16**, (2016).

19. U. Demšar, A. Çöltekin, Quantifying gaze and mouse interactions on spatial visual interfaces with a new movement analytics methodology. *PLoS ONE* **12**, (2017).

20. National Academy of Engineering, "Grand Challenges for Engineering" (National Academy of Sciences/National Academy of Engineering, Washington, DC, 2008).

21. B. P. Woolf, *Building intelligent interactive tutors: student-centered strategies for revolutionizing e-learning*. (Morgan Kaufman, 2008).

22. S. D. Craig, A. C. Graesser, J. Sullins, B. Gholson, Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media* **29**, 241 (2004).

23. G. Hatano, K. Inagaki, in *Annual Meeting of the American Education Research Association*. (New Orleans, 2000).

24. J. D. Bransford, D. L. Schwartz, in *Review of Research in Education,* A. Iran-Nejad, P. D. Pearson., Eds. (Washington, DC, American Educational Research Association, 1999), vol. 24.

25. C. W. Coultas, R. Grossman, E. Salas, in *Handbook of Human Factors and Ergonomics*. (John Wiley & Sons, Inc., 2012).

26. S. J. Pan, Q. Yang, A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345 (2010).

27. P. Durlach, R. Spain, in *Advances in Applied Human Modeling and Simulation,* V. G. Duffy, Ed. (CRC Press, 2012).

28. R. D. Pea, The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences* **13**, 423 (2004).

29. T. Murray, I. Arroyo, paper presented at the Proceedings of the 6th International Conference on Intelligent Tutoring Systems, 2002.

30. L. S. Vygotsky, *Mind and society: The development of higher psychological processes*. (Harvard University Press, Cambridge, MA, 1978).

31. J. A. Anderson, A. T. Corbett, K. Koedinger, R. Pelletier, Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences* **4**, 167 (1995).

32. P. Dillenbourg, J. Self, A framework for learner modeling. *Interactive Learning Environments* **2**, 111 (1992).

33. Z. A. Pardos, N. T. Heffernan, B. Anderson, C. L. Heffernan, in *Handbook of Educational Data Mining,* R. Christobal, Ed. (CRC Press, 2010), pp. 417-426.

34. J. T. Folsom-Kovarik, R. E. Wray, L. Hamel, paper presented at the Artificial Intelligence in Education (AIED), Memphis, 2013.

35. J. V. Cohn, D. Nicholson, D. Schmorrow, Eds., *The PSI Handbook of Virtual Environments for Training and Education*, vol. 3 (Praeger Security International, Westport, CT, 2008), vol. 3.

APPENDIX A: Interactions between Learner Assessment and Content Requirements: A Verification Approach

Interactions between Learner Assessment and Content Requirement: A Verification Approach

Robert E. Wray, Kimberly Stowers

1 Soar Technology, Inc. 3600 Green Court Suite 600, Ann Arbor, MI USA 48105
wray, kimberly.stowers @soartech.com

Abstract. A practical constraint in the design and development of algorithms and tools for personalized learning is the need to implement adaptive algorithms, oftentimes within complex software environments, without the benefit of a priori large-scale user testing. The lack of such testing makes it difficult to ensure that lessons and guidance from design recommendations and prior studies in other domains has been effectively applied in the training application. This paper summarizes efforts toward a testbed to support verification of adaptive training designs. The testbed operationalizes evidence-based guidance from the research literature and simulated students to enable exploration of design space prior to large-scale implementation. The paper motivates the approach with a specific design question, which is to examine trade-offs between the use of behavioral markers to assess proficiency and the resulting training-content requirements to take advantage of the information that such markers provide.

Keywords: training design, adaptive training

1 Introduction

A practical constraint in the design and development of algorithms and tools for personalized learning is the need to design, implement and integrate adaptive algorithms, oftentimes within complex software environments, without the benefit of a priori large-scale user testing. User testing can provide evidence of what adaptive methods are more (and less) beneficial within a particular training setting. The most beneficial, specific methods will usually not be fully known in advance; many potential design options may be apt. Knowledge of the research literature and results can be helpful, but best practices for the design of adaptive training in most training contexts is ever-evolving [1, 2].

This constraint is particularly acute in complex training environments, such as those used in distributed simulation and virtual training. The complexity of software integration and limited access to physical devices can result in commitment to a design that turns out to not offer many training benefits. Similarly, a chosen approach may offer a significant improvement in learning effectiveness but the target population cannot realize those benefits because their incoming knowledge and skill is not matched to those benefits provided by the system.

When an algorithm or approach turns out to be poorly chosen, it may take several years to develop and implement an alternative approach. This delay has both immediate and longer-term impacts. The immediate cost is the lack of improvements in training that were anticipated by the training developers. A longer-term, more systemic cost is that these failures in execution can impose greater resistance and new barriers for the adoption of adaptive training generally, resulting in the perception that adaptive training methods are not sufficiently mature to deliver the learning benefits that have been observed in more controlled (and, oftentimes, contained) settings.

As researchers interested in developing and fielding effective adaptive training solutions, we have for several years been developing a methodology that employs simulated students and software verification methods to attempt to understand the potential benefits of adaptive algorithms and the requirements they impose on students and instructors prior to full-scale development [3-5]. We introduce a testbed we are developing to enable exploration of design choices and, to illustrate how the testbed can inform specific design choices, summarize a verification study conducted using the methodology. This study reflects the long-term goal to develop methodology and tools that will help designers understand what (adaptive) features are appropriate/needed for their training needs and to estimate the costs/benefits of different design options.
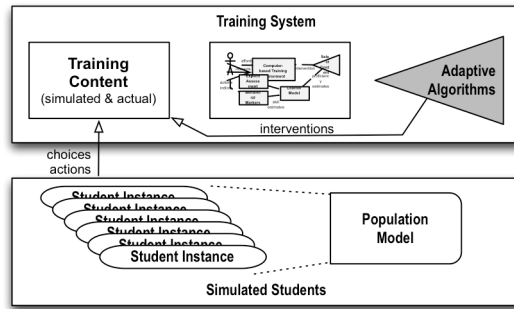

Fig. 1. Conceptual Architecture of Verification Testbed.

2   Testbed for Training Design

Below we briefly introduce the elements of the verification testbed we are developing. The goal of the testbed is to provide a computational tool, with parameters connected to the research literature, that allows a training designer to evaluate assumptions about a design. Fig. 1 illustrates the major components of the testbed and their relationships to one another.

Testbed components are:

1. **Adaptive algorithms**: The testbed typically uses the implementation of adaptive algorithms that would be used in the actual training environment. From a software engineering perspective, this approach allows evaluation and test (or *verification*) of the adaptive solutions within the testbed.

2. **Learning-system architecture**: The learning-system architecture defines how training content will be delivered and the role of adaptive algorithms within the learning environment. We are developing a family of these models for use in the

testbed. The next section introduces the specific model we are using for this analysis (see Fig. 2).

3. **Training content**: The testbed draws on a content repository to deliver training content within the testbed. In some cases, this training content may be the actual content that is to be used in the training application (especially apt when adding adaptive capabilities to an existing training application). In other cases, especially for a new training system being designed, the training content may be simulated.

4. **Simulated students**: The testbed employs simulations or models of students to interact with the training content. The use of simulated students to support training design is becoming more commonplace; some researchers have identified methods to synthesize functional students based on task analyses, cognitive architectures, and machine learning [6, 7]. Analytic tools, such as power law equations, are often also used for modeling learning [8, 9]. The primary requirement for a simulated student is that it provide a response to a learning situation at an appropriate level of abstraction for the simulation of the learning environment.

5. **Population Model**: The population model varies parameters for individual simulated students as they are instantiated. Having a distinct population model (rather than a defined population of simulated students) allows the user of the testbed to explore potential interactions between population assumptions (students with generally high/low self-efficacy; students generally well-prepared or poorly prepared for the content to be delivered).

Long-term, we envision a flexible and composable software environment that would allow designers to model potential learning designs and evaluate them in a decision analysis aid. Today, we are creating instances of the components illustrated in Fig. 1 to address specific design questions, as discussed next.

3   Motivating Example

As described above, the study we present uses a simulated students paradigm and a simulation of the learning environment to provide quantitative estimates for functional system requirements. The benefit of this approach is that specific learning benefits and the effects of adaptation can be evaluated, at least tentatively, in advance of full-scale implementation.

Fig. 2. Model of the learning environment.

Here we discuss the learning environment being simulated, along with the specific domain we pull learning content from.

Computer-based training (CBT) is actively used across many contexts, including military, medical, and educational. CBTs commonly include didactic instruction (text
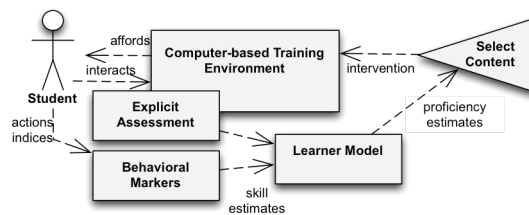
and images, audio, and video), opportunities for relatively simple practice, and periodic checks of knowledge. Most CBTs assume a fixed sequence of lessons and may require a student who fails a knowledge check to repeat a lesson. Implementing adaptive training in such a context may yield many benefits, most notably the benefit of accelerating or decelerating the pace at which students move forward in the lesson according to how quickly they are learning, including improved engagement. Adaptive techniques used in CBTs include variable starting points [10], enabling more/less practice [11], hinting and coaching [12, 13], and personalization of content delivery [14, 15].

We are designing and evaluating the role of adaptation in a CBT for Emergency Medical Technician (EMT) certification. EMT courses are offered across the United States, with various states enforcing slightly different requirements. Curriculum is standardized at the US federal level through the National Highway Traffic Safety Administration [16]. This makes EMT training both accessible and applicable. Additionally, EMT certification is a domain of training that can be applied in both national and international civilian and military contexts, making it a highly valuable area for the training improvement. Adaptive training may help streamline the EMT certification process by accommodating learners who may need more or less practice to meet national standards.

For the specific analysis of this paper, we examine a specific lesson in the standard curriculum for EMT training—scene size-up. Scene size-up involves steps taken by an EMT crew when arriving on the scene of an emergency. According to the standard curriculum, in order to develop training within this context, it is necessary to consider what a "scene size-up" timeline looks like, and cognitive, affective, and psychomotor objectives are for this task (see table 1). The standard curriculum specifies 9 distinct learning objectives across these three different types of learning objectives.

It would be useful in designing the training environment to have insights and quantitative estimates for the following three questions:

1. *What is the potential size of the learning gain that would be introduced by the use of adaptive methods?* This question sets expectations for the design and helps the designer to understand the relative benefit of adaptive training in the context of the impacts of the full system.

2. *How much unique content is needed to realize the ideal (or at least compelling) learning gains*? Tailoring to the learner typically requires specialized content. If we assume that it is not possible to automate content creation (the typical case), then it would be beneficial to estimate the minimum content needed to realize a (meaningful) gain from adaptive tailoring.

3. *How accurate do assessment measures need to be to realize (compelling) learning gains?* In order to make adaptive choices, some measurement of the state of the learner during the learning process is typically needed? How accurate do measures need to be to realize the hypothesized gains from adaptive tailoring?

Table 1. Key parameters for the marker/content verification analysis.

Parameter	Description	Study Value(s)	Citations
Base Learning Rate	The learning rate term in a standard power law learning curve (α)	.5	The specific α value is in the range of common values in learning models [8, 9]
Learning Objectives Types	Distinct categories of learning objectives.	3	Cognitive, Affective, Psychomotor from Standard EMT Curriculum [16].
Number of Learning Objectives	Objectives that must be met according to the topic and tasks being learned to complete a scene size-up.	9	9 distinct learning objectives are identified in the standard curriculum [16]
Z Score	A normalized (-1..1) relative match between learner capability and material being presented.	*See text*	This Z-score is an operationalization of the ZPD and is informed by [18] but is adapted to the anticipated training context.
Delta Learning rate	Modification of base learning rate with the assumption that high z-score improves learning rate and low z-score diminishes learning rate.	+/- 25%	This range is comparable to learning gains observed in a similar domain with tailored content matching [15].
Measure Accuracy	The general accuracy of measures used to estimate skill/proficiency.	*See text*	Direct measures can have high accuracy. Indirect measures, such as markers, often can exhibit poor precision and recall.

4. Verification Methodology

To attempt to answer these questions, we developed a simulation of the EMT learning environment within the testbed and developed specific tests to gather data. A summary of the implementation for each testbed component is summarized below. Table 1 lists specific values for some of the primary parameters used in the study. Testbed components:

1. **Adaptive algorithms**: This test focuses on a single adaptive algorithm, which chooses the lesson content that is closest to the estimated proficiency of the learner across all learning objectives. We are interested in the use of other adaptive algorithms, including hinting and coaching. However, in this study, we focus only on lesson selection.
2. **Learning-system architecture**: Modeled as displayed in Fig. 2. We did not distinguish explicit assessment and marker-based measurement, although explicit assessment is generally more accurate than marker-based techniques.
3. **Training content**: We generated several collections of lessons, which are primarily characterized by the target learner profile for the lessons (but not all lessons touch on all learning objectives). The comparison standard for lessons was

the "progressive" lesson design, which assumes an initial low student proficiency vector and increases the values in the profile across all learning objectives as lessons progress. This choice is reasonable for most CBTs, although a part-task design would be a contrasting option for future study.

4. **Simulated students**: In this design, students were simulated using a power law model. We employed a form of the power law model which computes the impact of a lesson solely from the current lesson and prior learning [17]. This form of the power law allows us to estimate the effect of each individual lesson and not assume a heterogeneous distribution of lessons. For the study, each "lesson" was estimated to be about 4 minutes of instruction, resulting in 15 distinct lessons (and 14 opportunities for intervention) within the learning design.

   The effect of adaption on learning is estimated by assessing how closely a chosen lesson matches the learner's proficiency profile. A Z(PD)-score is computed as the average mismatch between the lesson (target profile) and student/actual profile for all learning objectives addressed by the lesson. Normalization is applied to the average error to bound to the range [-1...1], where a 1 represents a perfect match and a -1 represents a (near-perfect) mismatch. How precise targeting needs to be is obviously of interest to the adaptive training community. We chose a conservative approach, assuming a functional relationship in which the maximum Z-score rapidly decreases for relatively small targeting errors. In other words, unless targeting is very good, its effect on learning rate will be small.

5. **Population Model**: The primary population variable used in the study is the initial proficiency profile of students. An initial proficiency profile for each student (100 students were generated per condition) was computed based on an initial bias (e.g., "very low", "low", "any") and a sampling of the normal distribution across that bias. Again, this approach does not yet account for students who may be more differentially prepared for the training (e.g., very low for some learning objectives, but high for others).

5 Results

We generated testbed simulations focused on the three questions introduced above. This section discusses a collection of tests, undertaken in the testbed, to help shine light on each question.

Fig. 3 summarizes one analysis of potential learning gains for Question 1. It illustrates hypothesized learning curves for two different populations. The "medium" initial proficiency populations (dotted lines) are assumed to have some prior knowledge/familiarity of the domain, resulting in an overall higher level of initial proficiency for the EMT Scene Size-up unit. For example, such students might already be able to recognize certain visual cues in a given scene such as broken glass or fuel spills and be familiar with relevant categorization terms (*trauma victim*) relative to scene size-up. The other population is assumed to have very low initial proficiency (dashed lines), meaning that they have little relative working knowledge of the EMT domain.

The figure compares learning rates for a well-designed curriculum (purplish lines) to those obtained using targeted content selection (blue lines). In these examples, we assume tailoring to the learner is accurate and that content can be tailored to each learner (unlimited content options). These conditions provide a "best case" difference between a well designed CBT and an adaptive one. The results of the analysis suggest that the benefit from adaptive content selection is likely to be relatively modest in comparison to a well-designed, progressive CBT. We expected to see greater separation for the learners with low initial proficiency, but the relative gains between the two populations are similar. In general, these results suggest that a training effectiveness/pilot study for this domain will be highly sensitive to the initial instructional design. Either more tailoring opportunities or more learning time may be needed to better separate adaptive and non-adapted learner populations.

Fig. 4 summarizes exploration of trade offs between adaptive tailoring and the con-
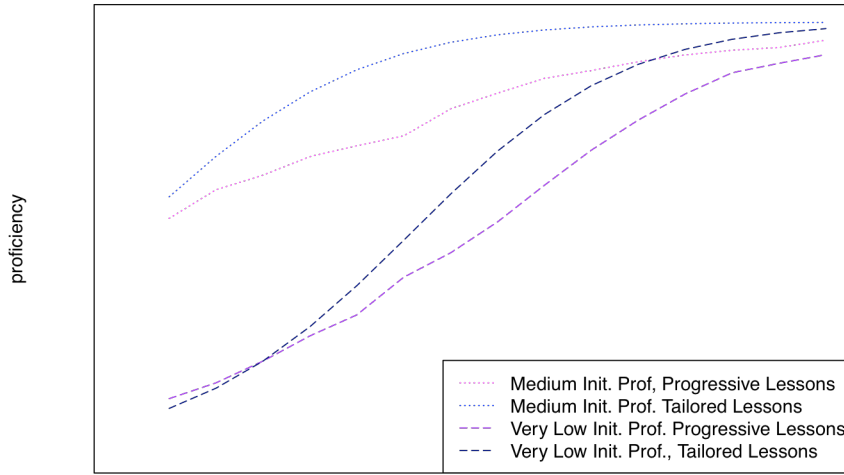
Fig. 3. Comparing Progressive (purple) & Tailored (blue) hypothesized learning trajectories for students with moderate a prior familiarity (dotted lines) and little familiarity (dashed).

tent available for adaptation. The figure contrasts projected learning outcomes under the same test conditions (other than available content) and uses the "very low" initial proficiency population as described for Fig. 3. The content options included in the figure are *unlimited* (content is available to match any proficiency profile) and a number of content choices: 2 choices (binary decision), 3-5 choices (small number of choices), and 10 choices (many choices). All choices were generated by sampling across the full spectrum of performance vectors. For example, for a 3 choice decision, one option would be generated for the "low", "medium", and "high" proficiency bias.

The figure suggests adaptive content selection is not likely to have a significant positive impact on learning unless sufficient content is available. Even 3-5 choices/decision were not sufficient to significantly improve learning. For continuing analysis, we plan to examine whether choices more localized to the typical learning progression (as reflected in the "progressive instructional design" in Fig. 3), could boost
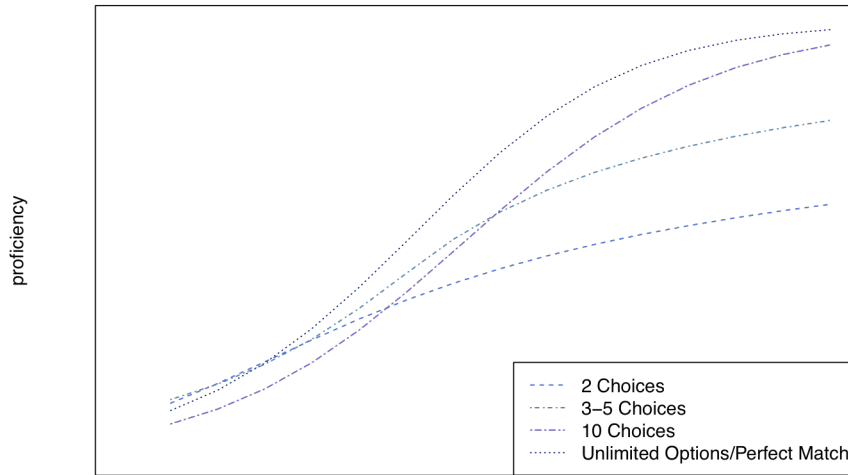
Fig. 5. The potential effects of content availability on learning outcomes.

the performance of adaptive content selection without requiring a prohibitive number of content options. In general, the worst-case performance for adaptive selection should be to just choose the choice in the original instructional design, so these results are somewhat more pessimistic than would be the case in actual implementation.

The final question was to attempt to quantify the accuracy of the underlying measures needed to enable adaptive tailoring. As shown in Fig. 2, we would like to use both explicit measures (e.g., a score from questions delivered after a lesson) as well as behavioral markers that provide (passive) indicators of learner state during learner activities in the CBT. Fig. 5 illustrates an initial assessment of the trade off inherent in using learner state measures to enable adaptive content selection. It presents learning curves obtained from a 95-70% range on measurement accuracy in
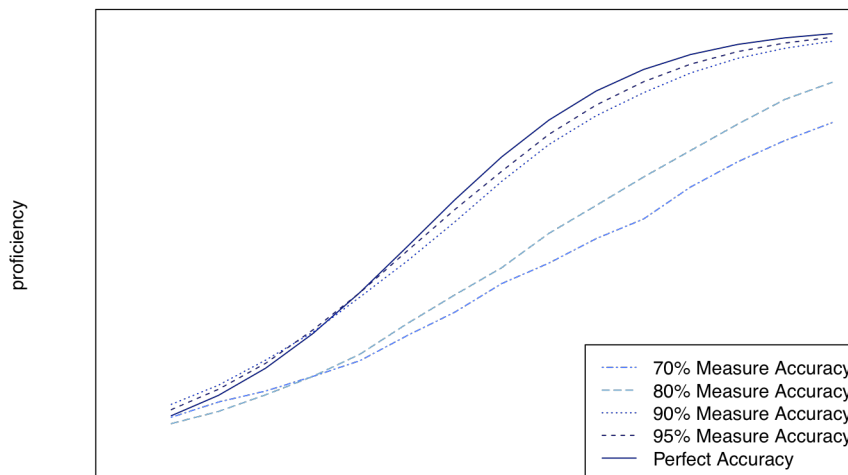
Fig. 4. The potential effects of measure accuracy on learning outcomes.

comparison to the learning curve obtained from perfect (100% accuracy) measures. Accuracy is computed as a normally distributed error around actual (ground-truth) levels of learner skill. It does not take into account compound errors across trials or reductions in measurement error with systematic, iterative measurement.

In general, as the accuracy of the measure degrades, the system's ability to narrow its tailoring to an individual learner's ZPD degrades as well. As suggested by the figure, even a (relatively good) 80% accuracy results in a loss of much of the advantage of adaptive content selection. This result, combined with the analysis summarized by Fig. 3, strongly suggests that adaptive content selection alone may not provide significant value for learning, given the limits of measurement accuracy, even if content requirement barriers could be mitigated (e.g., by some automatic content generation or content variation processes).

6   Conclusions

This paper illustrated an analytic approach to the design of adaptive training, enabling quantitative evaluation of design questions prior to commitments to implementation and pilot testing. In the illustrative example, analysis identified only marginal benefits of adaptive content selection in comparison to a well-designed learning environment. Further, realizing those small benefits requires unrealistic demands for accuracy in learner measurement and content creation. While these are somewhat negative results from of the point of view of advancing adaptive training, examples and tools supporting such analyses offer the potential to help researchers and practitioners set realistic expectations for learning system outcomes and to quantity component requirements within an adaptive training system to ensure minimum learning gains can be realized by an implemented system.

Acknowledgements

References

1.   Landsberg, C.R., Astwood, R.S., Van Buskirk, W.L., Townsend, L.N., Steinhauser, N.B., Mercado, A.D.: Review of adaptive training system techniques. Military Psychology 24, 96-113 (2012)
2.   Durlach, P.J., Lesgold, A.M. (eds.): Adaptive Technologies for Training and Education. Cambridge, New York (2012)

3.   Folsom-Kovarik, J.T., Wray, R.E., Hamel, L.: Adaptive Assessment in an Instructor-Mediated System. Artificial Intelligence in Education (AIED), (2013)

4.   Wray, R.E., Bachelor, B., Jones, R.M., Newton, C.: Bracketing human performance to support automation for workload reduction: A case study. LNCS: Proc. of HCII 2015 Conference. Springer-Verlag, Los Angeles (2015)

5.   Wray, R.E., Woods, A., Haley, J., Folsom-Kovarik, J.T.: Evaluating Instructor Configurability for Adaptive Training Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics (AHFE 2016) and the Affiliated Conferences. Springer, Orlando (2016)

6.   Matsuda, N., Cohen, W.W., Sewall, J., Lacerda, G., Koedinger, K.R.: Predicting students' performance with SimStudent that learns cognitive skills from observation. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) Proc. International Conf. on Artificial Intelligence in Education, pp. 467-476. IOS Press, Amsterdam (2007)

7.   MacLellan, C.J., Koedinger, K.R., Matsuda, N.: Authoring Tutors with SimStudent: An Evaluation of Efficiency and Model Quality. In: Proceedings of the 12th International Conference on Intelligent Tutoring Systems, pp. 551-560. (Year)

8.   Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S.A., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111, 1036-1060 (2004)

9.   Newell, A., Rosenblum, P.S.: Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (ed.) Cognitive Skills and Acquistion. Erlbaum, (1980)

10.   Eagle, M., Corbett, A., Stamper, J., McLaren, B.M., Wagner, A., MacLaren, B., Mitchell, A.: Estimating individual differences for student modeling in intelligent tutors from reading and pretest data. International Conference on Intelligent Tutoring Systems, pp. 133-143. Springer (2016)

11.   Lee, J.I., Brunskill, E.: The Impact on Individualizing Student Models on Necessary Practice Opportunities. Intrntl Educational Data Mining Society, (2012)

12.   Durlach, P.J., Ray, J.M.: Designing Adaptive Instructional Environments: Insights from Empirical Evidence. Army Research Institute for the Behavioral and Social Sciences (2011)

13.   Schatz, S., Oakes, C., Folsom-Kovarik, J.T., Dolletski-Lazar, R.: ITS + SBT: A Review of Operational Situated Tutors. Military Psychology, special issue on current trends in adaptive training for military ap-plication (2012)

14.   Lane, H.C., Johnson, W.L.: Intelligent Tutoring and Pedagogical Experience Manipulation in Virtual Learning Environments. In: Cohn, J., Nicholson, D., Schmorrow, D. (eds.) The PSI Handbook of Virtual Environments for Training and Education, vol. 3. Praeger Security International, Westport, CT (2008)

15.   Chaplot, D.S., Rhim, E., Kim, J.: Personalized Adaptive Learning Using Neural Networks. In: Proc. 3rd ACM Conference on Learning @ Scale. (2016)

16.   United States Department of Transportation, National Highway Traffic Safety Administration: EMT-Basic: National Standard Curriculum. (1996)

17.   Leibowitz, N., Baum, B., Enden, G., Karniel, A.: The exponential learning equation as a function of successful trials results in sigmoid performance. Journal of Mathematical Psychology 54, 338-340 (2010)

18.   Murray, T., Arroyo, I.: Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. Proceedings of the 6th International Conference on Intelligent Tutoring Systems, pp. 749-758. Springer-Verlag (2002)

APPENDIX B: PROTOCOL FOR HUMAN SUBJECTS STUDY

Assessing the Role of Behavioral Markers in Adaptive Learning
Study Protocol
Version 1.5.8 (8 May 2017)

Robert E. Wray, PhD (PI)
Soar Technology, Inc.

APPENDIX A: DEMOGRAPHIC QUESTIONNAIRE (ATTACHED)

APPENDIX B: LEARNING ENVIRONMENT TUTORIAL & MARKER CALIBRATION (ATTACHED)

APPENDIX C: EXAMPLES OF SUBJECT PRE-/POST-TEST QUESTIONS (ATTACHED)

APPENDIX D: EXAMPLES OF INSTRUCTIONAL CONTENT PRESENTED TO SUBJECTS (ATTACHED)

APPENDIX E: RECRUITMENT FLYER

A. Introduction and Background

Personalized learning, in which a learning environment adapts to the abilities, needs, and preferences of individual learners, has been identified as a "Grand Challenge" for 21st century research and engineering (National Academy of Engineering, 2008). The benefits of adaptive learning environments include more efficient learning (Woolf, 2008), improved attention and motivation (Craig et al., 2004), the development of less rigid and more flexible decision making (i.e., adaptive expertise, Hatano & Inagaki, 1986), and improved transfer of learning to settings in which learned knowledge is used and applied (Bransford & Schwartz, 1999; Coultas, Grossman, & Salas, 2012; Pan & Yang, 2010). Improved and personalized learning has particular application for more pervasive and less costly medical training, which often is delivered primarily by human instructors in classes with modest student-to-teacher ratios. Human instruction and mentoring is very valuable and desirable, but adaptive personalization methods offer an opportunity to deliver good, effective introductory and basic training, thus potentially enabling a single human instructor to train many more students by better preparing them for coaching and instruction from experts.

Adaptation to a learner usually requires a model of the learner that is frequently updated as a learner progresses through a curriculum (Durlach & Spain, 2012). The targeting of adaptive techniques, such as scaffolding (Pea, 2004) and competency matching (Murray & Arroyo, 2002; Vygotsky, 1978), depends on the accuracy (and, to some degree, precision) of the learner model. When the model better reflects the learner's actual knowledge, skills, and attitudes at any point during the learning, the targeting of the adaptive method to the learner generally improves (Murray & Arroyo, 2002).

Creating a complete and accurate learner model is difficult, however. In addition to estimating learner capability from formal and informal assessment within the environment (Anderson et al., 1995; Dillenbourg & Self, 1992; Durlach & Spain, 2012; Pardos et al., 2010), researchers have explored many behavioral, physiological, and even neurological indicators or "markers" that can provide additional context for estimating a learner's cognitive state and improving the dynamic assessment of the learner . For example, behavioral sensors (posture, eye trackers), physiological sensors (Galvanic skin response), and neurological sensors (EEG) have all been used to assess and track learner arousal/attention in learning environments (Cohn, Nicholson, & Schmorrow, 2008). Further, understanding the dynamic patterns of learner attention/arousal allows the identification of dynamic adaptation targeted to the identified arousal states (Cohn, Kruse, & Stripling, 2005).

Such markers can be useful for improving a learner model, but most markers today require sensors that are not commonly available on the hardware available for typical computer-based learning: a laptop or a tablet. The primary goal of this study is to assess the role of behavioral markers that have the potential to improve learner modeling while also not requiring specialized hardware/sensors (i.e., using only hardware sensors found on typical computing devices). The study focuses specifically on behavioral markers that can be derived from 1) mouse movements and mouse selections ("clicks") and 2) patterns of eye movements observable from a web camera ("passive eye tracking").

There is significant and growing scientific evidence that the temporal patterns of mouse movements during selection tasks can provide reliable insight into the cognitive state of subjects (Hehman, Stolier, & Freeman, 2015; Quétard et al., 2016). We anticipate, however, these markers to be noisier (less diagnostically precise) than neuro-cognitive markers associated with specialized sensors. Thus, this study focuses on evaluating the impact of the behavioral markers on the adaptive learning system to improve learning outcomes, given the noise and uncertainty of measure inherent in these unspecialized sources.

Under this study, multiple hypotheses will be explored:

- H1: There is a difference between conditions such that learning outcomes from the adaptive condition will exceed those from the non-adaptive condition.
- H2: Mouse movements will be an indicator of learner focus on certain aspects of the learning environment.
- H3: Eye movements will be an indicator of learner focus on certain aspects of the learning environment.
- H4: Mouse and eye movements will be correlated.

B. Study Design

In order to explore the hypotheses discussed in section A, a research study will be implemented which compares the results of learning between an adaptive medical learning unit to a unit presented in a non-adaptive (fixed) sequence. Specifically, curriculum units will be developed for "Scene Size Up," a required curriculum component used in Emergency Medical Technician (EMT) training (United States Department of Transportation & National Highway Traffic Safety Administration, 1996). These units (both adaptive and non-adaptive) will be presented to university subject population(s) in order to assess the utility of markers to improve adaptive learning in emergency medical environments. As discussed in section E, we will use multiple routes of recruitment, which will allow us to complete the study between July 1st, 2017 and January 31st, 2018.

Specifically, the following variables of interest will be implemented and observed:

- **Instructional approach**: The overall instructional approach of the learning environment. For this study, there are two distinct instructional approaches:
  - **Non-adaptive/traditional**: An instructional unit that is presented in a fixed sequence to all learners.
  - **Adaptive based on performance (only)**: An instructional unit in which specific content presentations are constructed/chosen based on learner performance and subsequent estimates of learner knowledge and skill.
  - **Adaptive based on performance and markers**: An instructional unit that is dynamically constructed/chosen based on a combination of direct learner observation (as above) and behavior markers.
- **Markers**: Patterns of observed behavior that are hypothesized to have a role in improving a learner model.
- **Knowledge gain**: A measure of the post-test performance of subjects, relative to pre-test performance.

This study will be implemented as a between-subjects design, with "instructional approach" being the independent variable of interest. Instructional approach will be manipulated at three levels (as discussed above): non-adaptive, adaptive based on performance (only) and adaptive based on performance and markers. To maintain the integrity of results, assignment will be randomized, with neither participants nor the experimenter being aware of assignment ahead of time.

Primary Experimental Conditions

Non-Adaptive (Standard Presentation)

Adaptation (Performance)

Adaptation (Performance and Markers)

The primary dependent variable will be "knowledge gain", as measured by difference scores between pre- and post-tests given to participants. Additionally, the behavioral markers outlined in section A, derived from dynamic tracking of mouse movements and eye movements, will be used to predict learner needs and adapt the learning environment. The combination of these variables will enable the study to address the hypotheses above, as well as quantify the utility of the chosen adaptive learning models for improving learning in medical environments.

C. Procedure

The procedure implemented for participants in this study is expected to take between 45 and 75 minutes. Specific steps in the procedure are detailed chronologically below.

1. Upon arrival, participants will read and sign the informed consent document.
2. Once participants have indicated their consent, they will be randomly assigned one of the three experimental conditions.
3. All participants will be given a standard demographics questionnaire (Appendix A) to assess their education level and familiarity (if any) with EMT training or medicine.
4. All participants will receive a short 5-minute tutorial on how to use APACTS (see Appendix B).
5. Passive eye tracking and mouse tracking mechanisms will be calibrated during the tutorial. Calibration includes the following standard practices:
    1. For eye tracking, adjustment of cameras and gaze calibration will be completed. This will require minimal activity from the participant, such as being asked to look around the screen (see Appendix B for example).
    2. For mouse tracking, calibration of the mouse will be completed. This will require minimal activity from the participant, such as being asked to move the mouse around the screen (see Appendix B for example).
6. All participants will complete a pre-test, developed by the experimenters, which contains questions about the process of completing the scene size-up task as an EMT (see Appendix C).
7. In their assigned condition, participants will learn how to complete a scene size-up, which will include the following standard practices for EMT training (see Appendix D for example content):
    1. Learning scene size-up terms and associated tasks.
    2. Viewing images of emergency scenes and reading text-based descriptions of the emergency scenes viewed.
    3. Viewing images of emergency scenes with opportunities to practice concepts learned, such as answering a question or labeling areas in a displayed image.
8. During their completion of these conditions, passive eye tracking and mouse tracking will be engaged to collect participant data.
    1. In the adaptive conditions, results from passive eye tracking and mouse tracking will be used to change what content is presented to the learner, such as varying the difficulty of practice tasks, presenting feedback customized to a subject's response, and/or repeating or amplifying previously presented information.

2. In the non-adaptive condition, the content presentation will not differ; all subjects will receive the same information, with identical feedback and level of difficulty as all other subjects.
9. During completion of conditions, participants will also receive questions tracking their sense of progress / self-efficacy in the domain.
10. Participants will complete a post-test, which will be identical to the pre-test (Appendix C).
11. Participants will be given an opportunity to give verbal feedback about the study before they leave.

D. Inclusions / Exclusion Criteria

The following inclusion/exclusion criterion will be adhered to and verified for each participant:

- Must be 18+ years old

The primary population of subjects will be college students, due to the source of recruitment (detailed in section E). College students represent a apt population for studying professional (in this case EMT) training, as they are pursuing professional endeavors that require similar training and learning practices. At the same time, the principle of distributive justice applies in this context, as college students represent a low risk population that can benefit from participation in research (through class credit or payment; see section E), and the study research is likewise low risk.

E. Recruitment of Participants

Primary Study Site: University of Alabama

The primary source of participants is the University of Alabama. Participants will be recruited from the University of Alabama through 3 different methods:

- Volunteers from University of Alabama's GBA300 classes, who are able to receive class credit for participation.
- Volunteers from University of Alabama's research participant pools, including Psychology Sona and CCIS participant pool, which are used to grant class credits.
- Paid participants recruited through flyers posted through University of Alabama's campus and on social media websites (see Appendix E).

Recruitment will begin in August 2017, with flyers/announcements being posted in classes and listed in the participant pools (per above list). We will not be requesting a set number of participants from each source. Instead, participants will be recruited freely through the above methods until the required sample size is met (see section I). Recruitment will be performed by the sub-investigator on the project, who has CITI certification through completing the "Group 2: Social Behavioral and Education Research Investigators and Key Personnel" course.

Secondary Study Site: Soar Technology, Inc. (Orlando Office)

Some subjects, especially for initial system testing and pilot assessment, will be recruited from the University of Central Florida (UCF) and Research Park areas. These subjects will exclusively be paid participants recruited through flyers posted through UCF's campus, Research Park (adjacent to UCF), as

well as email and social media websites (see flyer in Appendix E). Recruitment will be coordinated by both the Principal Investigator (Wray) and the sub-investigator (Stowers). Both have CITI certification. Subjects recruited at UCF will complete the study at the Orlando offices of Soar Technology, which is located in Research Park. An office will be dedicated for data collection at Soar Technology.

F. Consent Process and Timing

Consent will be obtained upon participant arrival to the research site. Before beginning the study, participants will be given a copy of the informed consent to read (the consent form will be developed by E&I for this study and thus is not attached to this submission). The experimenter will also explain the consent to them verbally. Participants will be given as much time as they need to consider participation and will consent verbally, as well as through written signature, before proceeding with the study.

The consent process will be performed the PI, the sub investigator and research assistants. All experimenters will have CITI "Group 2: Social Behavioral and Education Research Investigators and Key Personnel" certification.

G. Risks, Discomforts, and Benefits to Subjects

Minimization of Risks

Due to the nature of content used in the study, participants may find some of the images in the study disturbing (accident victims). These risks will be minimized through the use of images that minimize the visible presentation of injuries.

Maximization of Benefits

Participants will learn how to assess a medical emergency, and may find that learning process intrinsically rewarding. Benefits will be maximized through the use of practice rounds, as well as pre-tests and post-tests, where participants will be able to demonstrate their success in learning the content presented.

Provisions to protect the privacy of participants:

Privacy of Participants and Confidentiality of Data

Participant information will only be identified through assigned identification numbers. Through the use of the identification numbers, the data will be fully anonymous. Information connecting identification numbers with any personally identifiable information will be held in a separate location from other data collected and stored on a password protected computer. Only those involved in the study will have access to any information or data linked to the study.

Data Storage

Data will be stored for 5 years, according to guidelines by CITI. Data will be stored on a password-protected computer at all times and only the principal investigator and sub-investigator will have access to individual data.

H. Financial Considerations

Participants will be compensated $15 for participation via a credit-card gift card. Compensation will be provided at the end of the experimental session. Participants are not expected to incur any costs to themselves as a result of participation. If any research related injuries are discovered, the principal investigator and IRB will be notified immediately, as well as the University of Alabama's counseling and medical centers. Participants will have direct access to health care and counseling as needed.

I. Data Analysis and Statistical Analysis

As this study involves a single independent variable with just three levels, the primary analysis will be an F test comparing the difference scores of pre- and post-tests in each condition. Additionally, correlations will be calculated in order to gain an understanding of the relationship between behavioral markers and performance outcomes. A power analysis was run (using GPower 3.1) based on the following criteria:

- F test (one-way ANOVA)
- Effect size (*f*): 0.4
- Error probability (*alpha*): 0.05
- Power (*1 - beta error probability*): 0.85
- Number of groups: 3

According to the parameters entered and calculations made using GPower, we will need to analyze data from 72 participants to achieve optimal power. In order to account for participant withdrawal, as well as any issues encountered with eye tracking or mouse tracking that may cause data to be unusable (e.g., an adaptive condition in which mouse tracking did not function), we will collect data from up to 100 participants.

Analyses of participant data will be broken up into the following steps, the final step marking the endpoint of the study:

1. Coding and cleaning mouse-tracking and eye-tracking data
2. Calculating difference scores for pre- and post-tests
3. Calculating t-test and correlations
4. Reporting results through technical reports and publications

Our expectation is that all primary data analysis will be concluded by April 30, 2018. However, as data will be kept up to 5 years past the end of collection (see section G), we expect to also analyze depersonalized data on an ongoing basis. In particular, we will data captured from eye tracking and mouse tracking to inform further development and refinement of the markers tested in this study. For example, we are focusing a single mouse-tracking algorithm for use in the study. After the study is completed, we can perform post-hoc analysis with participant mouse tracking data to evaluate alternative mouse tracking algorithms and possible pattern-based selection of algorithms for future studies. Thus, the data resulting from this experiment will support subsequent research and improvement of adaptive learning methods and tools.

References

Anderson, J. A., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A Simple Proposal With Multiple Implications. In A. Iran-Nejad & P. D. Pearson. (Eds.), *Review of Research in Education* (Vol. 24). American Educational Research Association: Washington, DC.

Cohn, J. V., Kruse, A., & Stripling, R. (2005). Investigating the transition from Novice to expert in a virtual training environment using neuro-cognitive measures. In D. Schmorrow (Ed.), *Foundations of Augmented Cognition*. Mattawan, NJ: LEA.

Cohn, J. V., Nicholson, D., & Schmorrow, D. (Eds.). (2008). *The PSI Handbook of Virtual Environments for Training and Education* (Vol. 3). Westport, CT: Praeger Security International.

Coultas, C. W., Grossman, R., & Salas, E. (2012). Design, Delivery, Evaluation, and Transfer of Training Systems *Handbook of Human Factors and Ergonomics*: John Wiley & Sons, Inc.

Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media, 29*(3), 241-250.

Dillenbourg, P., & Self, J. (1992). A framework for learner modeling. *Interactive Learning Environments, 2*(2), 111-137.

Durlach, P., & Spain, R. (2012). Framework for Instructional Technology. In V. G. Duffy (Ed.), *Advances in Applied Human Modeling and Simulation*: CRC Press.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Asuma & K. Hakauta (Eds.), *Child Development and Education in Japan* (pp. 262-272). San Francisco: Freeman.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations, 18*(3), 384-401.

Murray, T., & Arroyo, I. (2002). *Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems*. Paper presented at the Proceedings of the 6th International Conference on Intelligent Tutoring Systems.

National Academy of Engineering. (2008). *Grand Challenges for Engineering*. Washington, DC: National Academy of Sciences/National Academy of Engineering.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345-1359.

Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2010). Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks. In R. Christobal (Ed.), *Handbook of Educational Data Mining* (pp. 417-426): CRC Press.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423-451.

Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., et al. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision, 16*(11).

United States Department of Transportation, & National Highway Traffic Safety Administration. (1996). EMT-Basic: National Standard Curriculum.

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Woolf, B. P. (2008). *Building intelligent interactive tutors: student-centered strategies for revolutionizing e-learning*: Morgan Kaufman.

Appendix A

Demographics Questionnaire

1. How old are you?

- __ (Fill in the blank)

2. Are you male or female?

- male
- female
- other

3. What is your education level?

- Graduated high school
- Completed some college coursework
- Completed Associate's degree
- Completed Bachelor's degree
- Completed Master's degree
- Completed Doctoral degree
- Other (please explain)
    - __ (fill in blank)

4. What is your major of study?

- __ (Fill in the blank)

5. Do you have any training or experience as an emergency medical technician or related service?

- Yes
- No

6. Do you have any formal training in first-aid procedures (such as a CPR course or training as a lifeguard)?

- Yes
- No

7. If yes to Question 5 or 6, please sketch some details (what training, when, etc.).

- __ (Fill in the blank)

Appendix B

Environment Tutorial & Calibration

Standard tutorial introduction to the instructional content delivery system (APACTS)

APACTS supports embedded videos

The "Coach" is used to provide directions, amplifying information, additional explanation, etc.

Introducing a "choice frame" (multiple choice questions)

Introducing annotation frames (tag locations within an image)
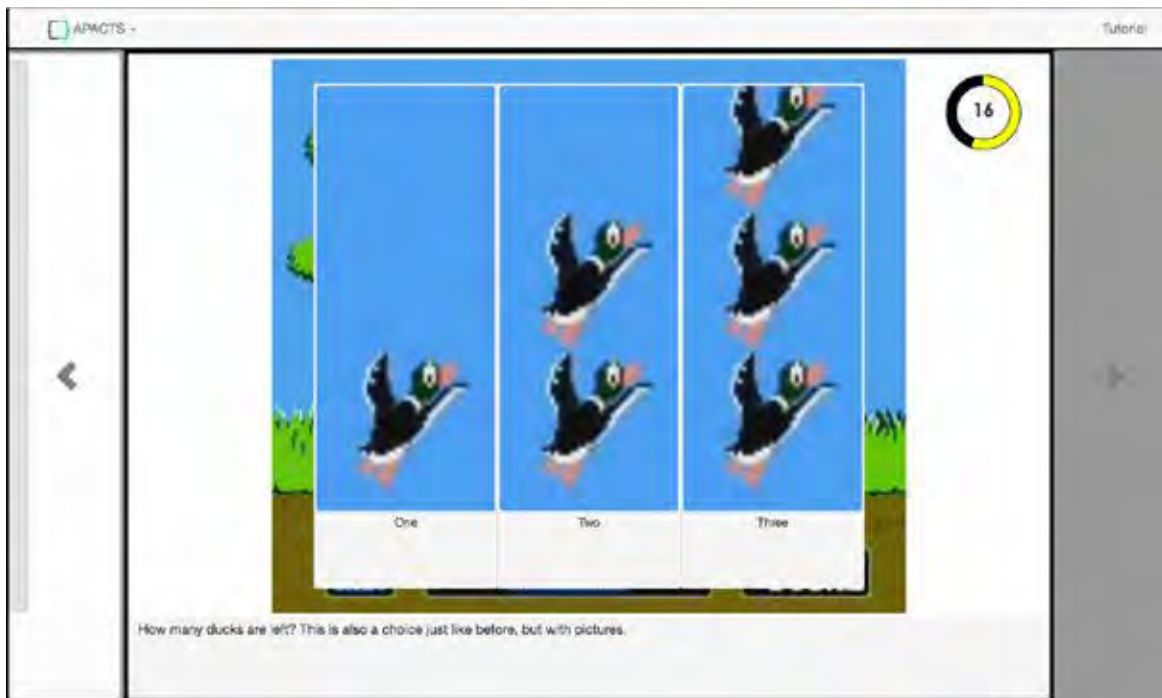
Choice frames can include images and text.

An alternative annotation frame

Calibration pattern.
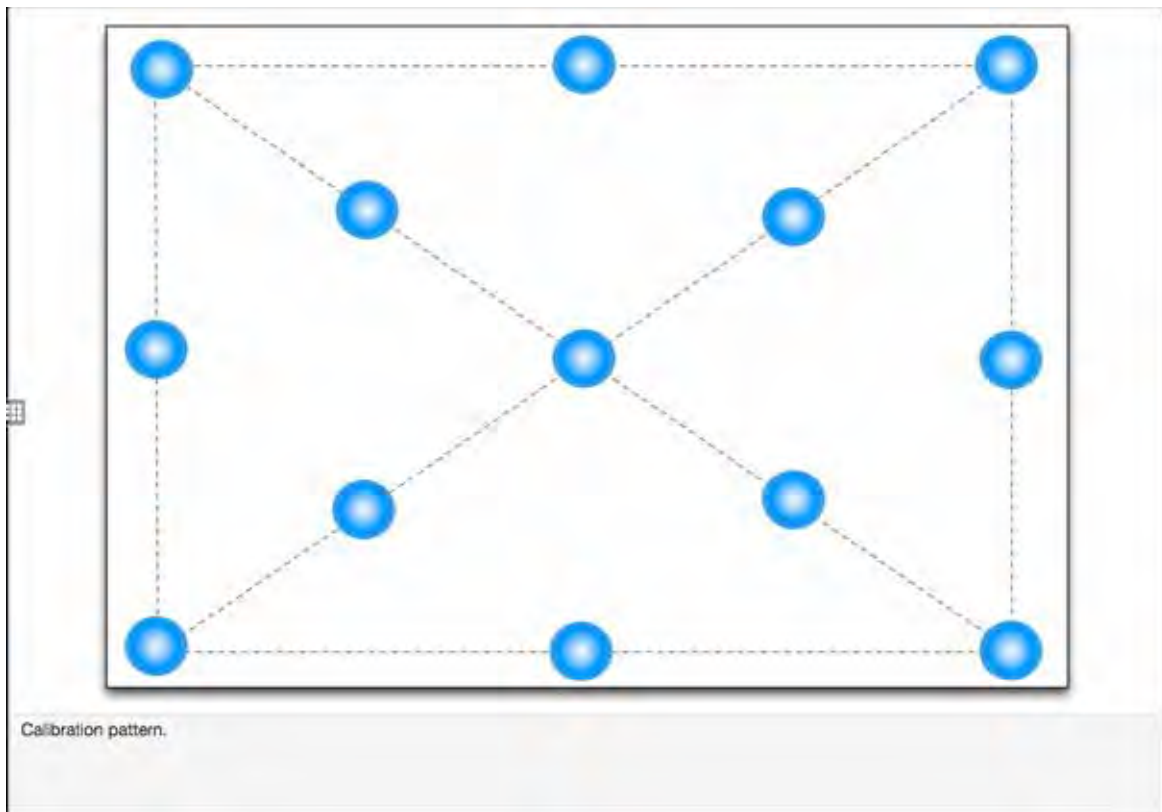
The tutorial will include simple calibration patterns for eye and mouse movements
(This image shows the underlying calibration pattern.)

Look at the blue ball when it appears on the screen.
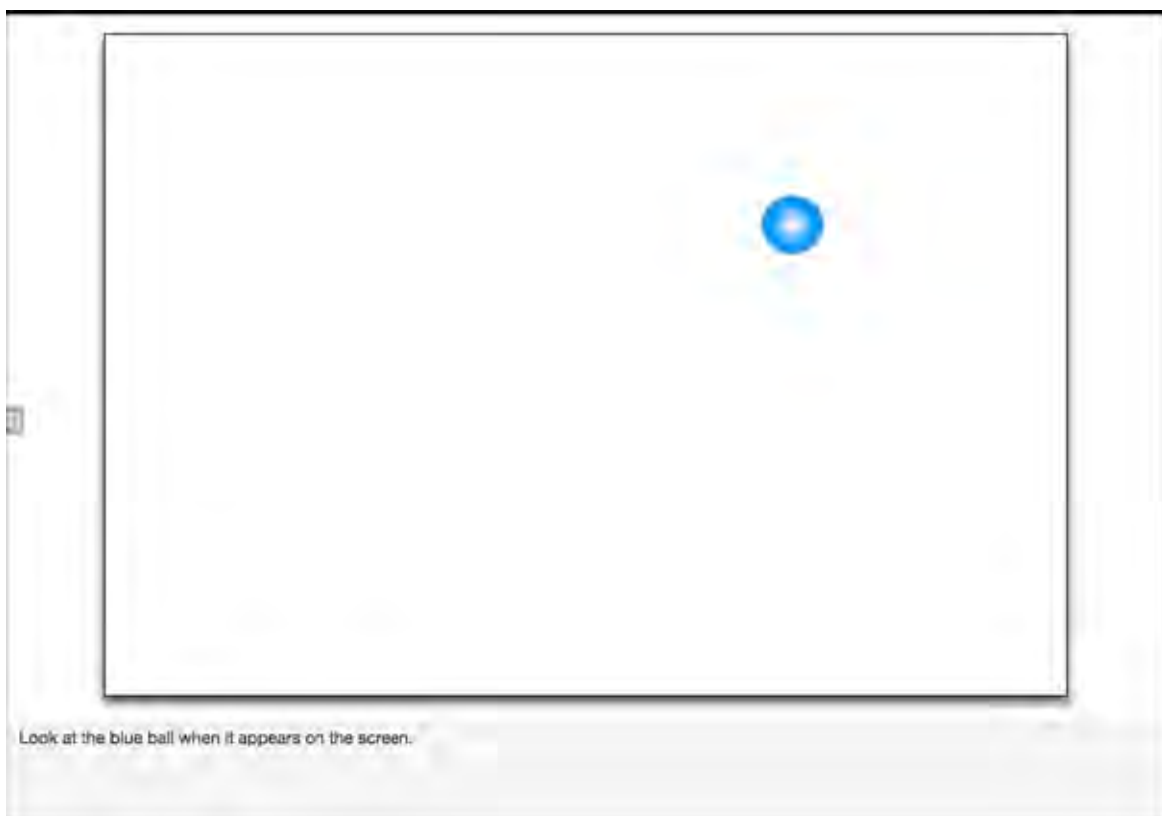
The actual calibration task will be 1) to fixate on a series of screen locations based on pattern, …

Move the mouse to the blue ball when it appears. Click when the mouse reaches the center of the ball.
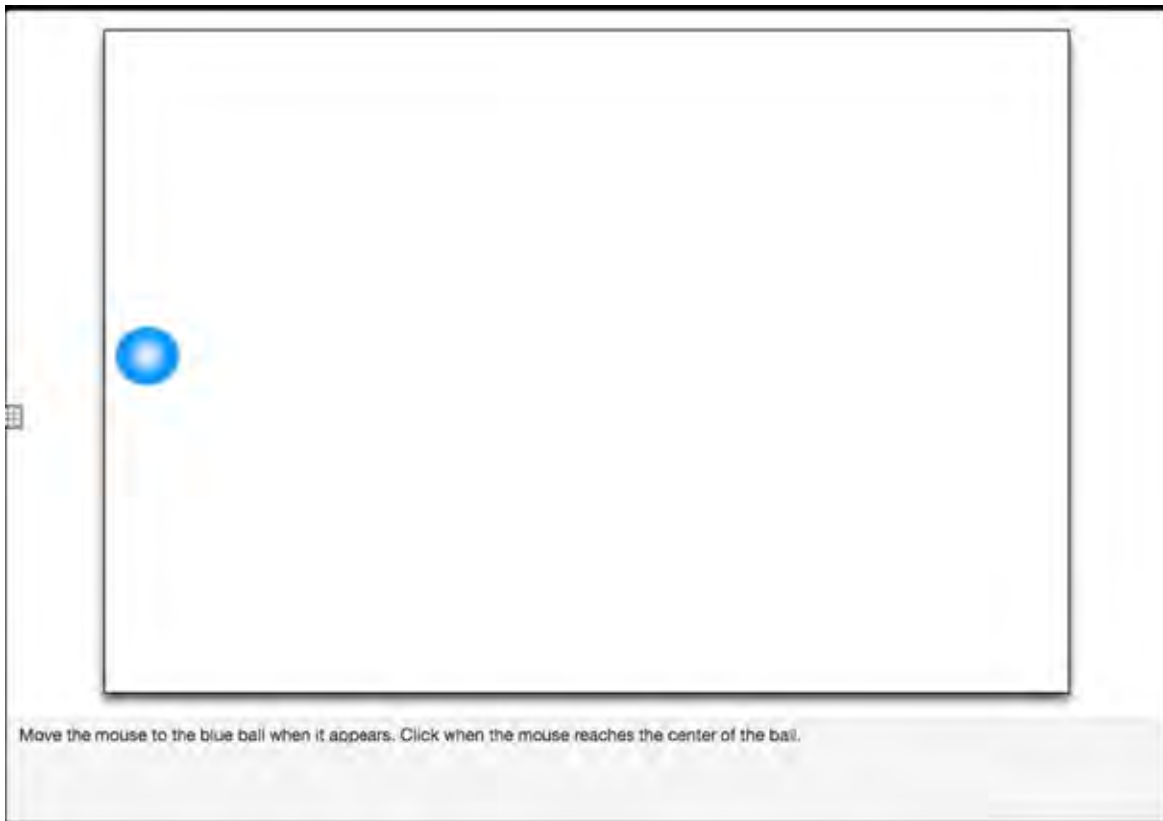
And 2) to move the mouse to a subsequence series of screen locations.

Appendix C. Pre-Test/Post-Test Example Questions

Subjects will complete a pre-test and post-test as part of the study. The pre-test and post-test will both be administered within the computer-based learning environment in which learning content is delivered (see Appendix D for specific examples of how questions are delivered within the system).

The pre-test and post-test will be identical and will not include any adaptive choices (the specific questions and their order will be fixed for all subjects/experimental conditions).

Below, we provide examples of the pre-/post-test questions for the study.

Basic Conceptual Knowledge

1. Which of the following best expresses the definition of *mechanism of injury* (MOI)?
    (a) The types of injuries observed for particular kinds of accidents
    (b) The immediate cause(s) of an injury that results from an accident
    (c) Mechanical failures in a vehicle (e.g., a blow out) that result in accident and injury
    (d) Action(s) that lead to accident and injury (failure to yield)
    (e) Both (c) and (d)

2. Which option best describes when scene size-up should be undertaken?
    (a) As soon as possible after arrival, but after immediate patient triage
    (b) During transit to the accident location, as provided by emergency personnel on scene via radio (or similar)
    (c) Immediately on arrival
    (d) After hazards have been assessed and bystanders moved away from hazards
    (e) Both (a) and (d)

3. What patterns of injuries are associated with side-impact collisions?
    (a) Head and neck injuries
    (b) Knee, hip, and leg injuries
    (c) Direct, blunt trauma
    (d) Broken arms and ribs
    (e) Both (a) and (b)
    (f) Both (a) and (c)
    (g) (a), (b) and (c)

4. What pattern(s) of injury are most associated with the "Down and Under" mechanism of injury?
    (a) Head and neck injuries
    (b) Knee, hip, and leg injuries
    (c) Direct, blunt trauma
    (d) Broken arms and ribs
    (e) Both (a) and (b)
    (f) Both (a) and (c)

4. What pattern(s) of injury are most associated with a roll over mechanism of injury?
    (a) Head and neck injuries
    (b) Knee, hip, and leg injuries
    (c) Direct, blunt trauma

    (d) Broken arms and ribs
    (e) Both (a) and (b)
    (f) Both (a) and (c)
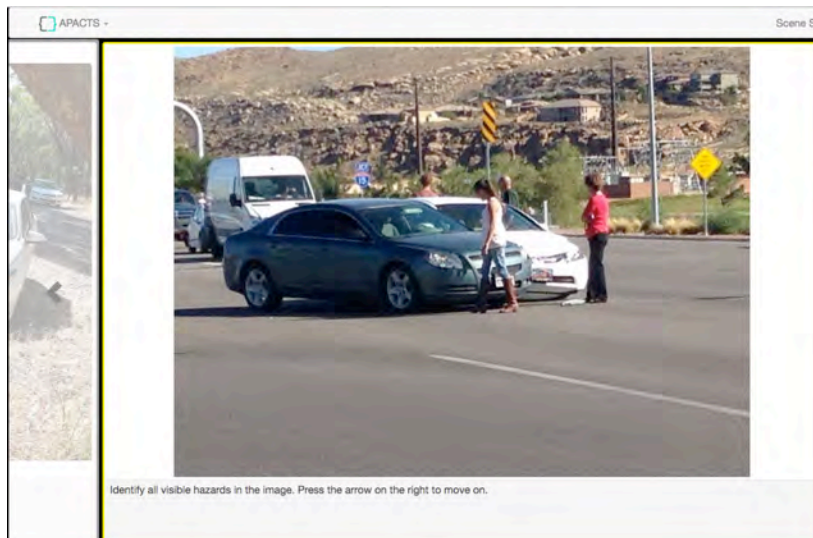   (g) All of the above

In addition to general knowledge questions, the pre- and post-test will include questions that present an image of an accident and ask the subject to evaluate the situation (size up the scene) in accordance with materials presented in the learning unit. These questions will be similar to the assessment and feedback questions that are used within the learning environment (i.e., as summarized in Appendix D).

Examples:

Application to a specific situation (multiple choice)
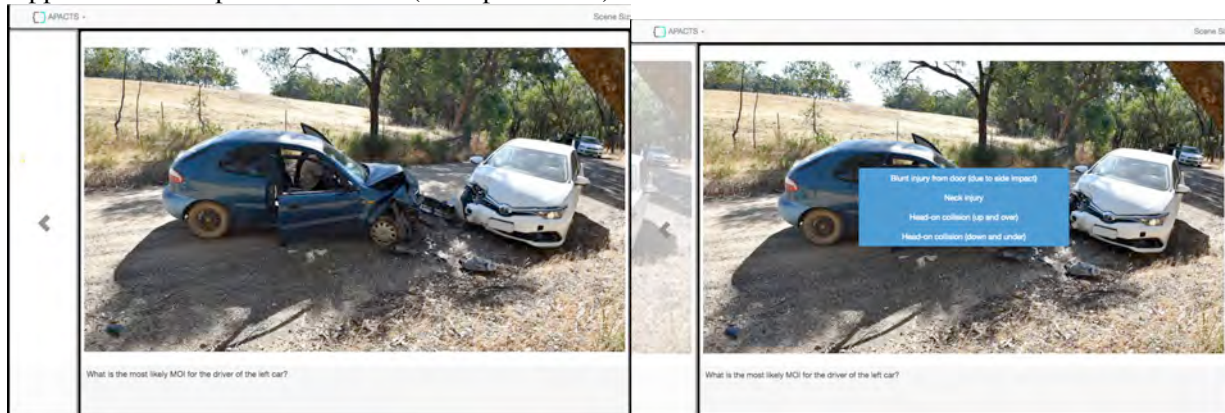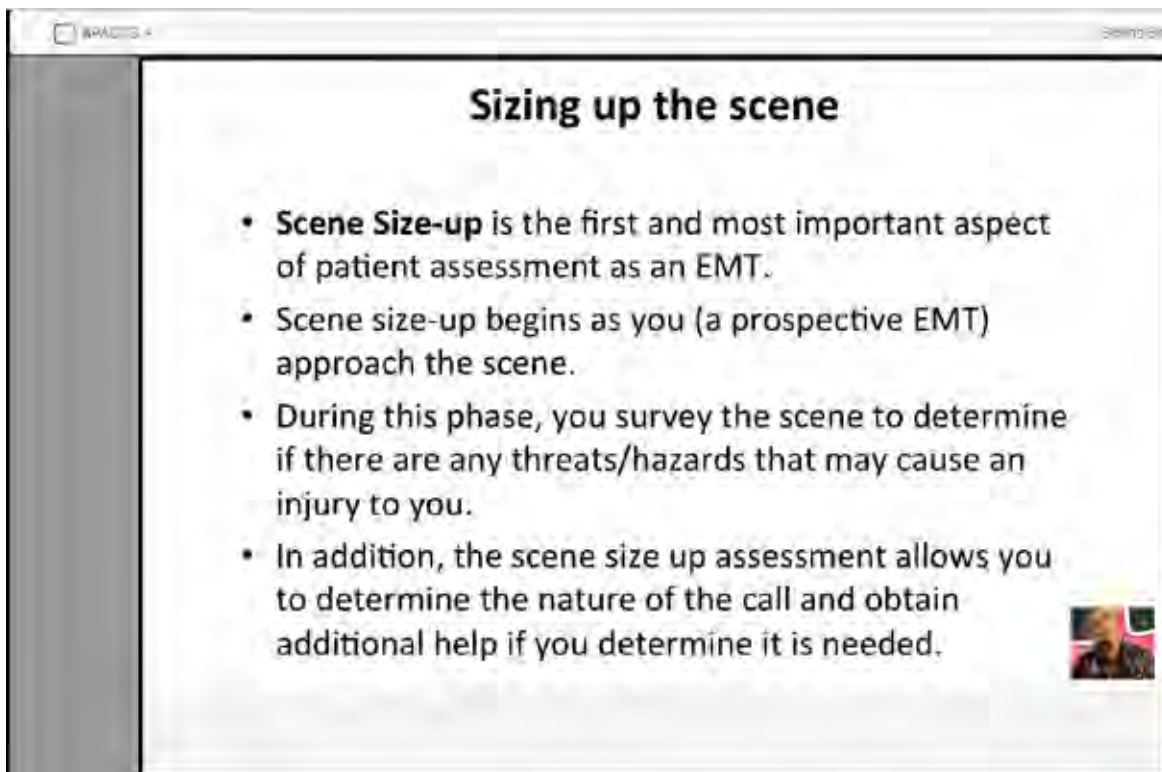
Application to a specific situation (labeling/annotation)

Appendix D

Example Content from the Learning Environment

Introductory instructional material

Objectives

During this lesson, you will learn to:

- Recognize hazards/potential hazards at a scene
- Describe common hazards found at the scene of a trauma (e.g., a car accident)
- Determine if a scene is safe to enter
- Anticipate common mechanisms of injury for several different kinds of car accidents
- Understand the reason for identifying the total number of patients at the scene
- Explain the reason for identifying the need for additional help or assistance
- Observe various scenarios and identify hazards

Ready to get started?

There's a lot here but you will have lots of opportunity to practice and receive feedback as we go. What do you expect to be hazards on arriving at a car accident?

Objectives of the unit of study. Clicking on the "coach" will bring up amplifying or summary statements.

Scene Safety

- Assess the scene to assure your well-being
- Personal protection - Is it safe to approach the patient?
  - Crash/rescue scenes
    - Fuel spills, broken glass, surrounding traffic
  - Toxic substances
  - Crime scenes - potential for violence; presence of guns
  - Unstable surfaces: slope, ice, water
  - Animals
- Protection of the patient - environmental considerations
- Protection of bystanders - if appropriate, help the bystander avoid becoming a patient.
- If the scene is unsafe, make it safe. Otherwise, do not enter

More detailed lesson material.

Do you see any hazards in this image? What might endanger you or accident victims?
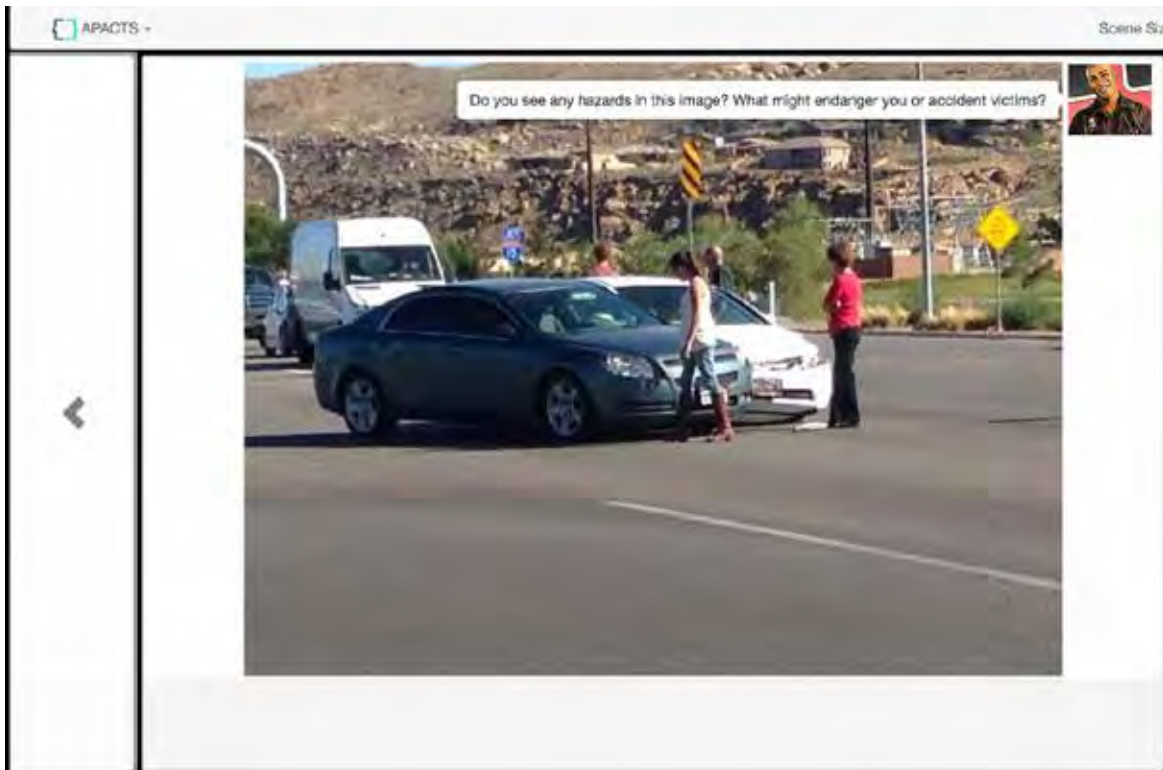
Opportunity to anticipate and consider more detailed explanation.

Scene Safety Questions

- Is it safe to approach the patient?
    - Look/listen for other emergency vehicles
    - Is there traffic? Is the flow safe?
    - Is there fire or smoke?
    - Look for hazardous materials (including debris from the accident)?
- Could this be a crime scene?
    - Fighting or loud voices?
    - Visible weapons?
    - Visible evidence of drug/alcohol use?
    - Unusual silence?
- Other factors
    - Pets can be a danger. Even friendly looking dogs could attack if they feel threatened.

"Check your knowledge" questions. Responses to these questions are used to update the learner model and influence subsequent content choices.

Simulated user response…

User receives feedback based on their response (both traffic and debris are hazards in the image).
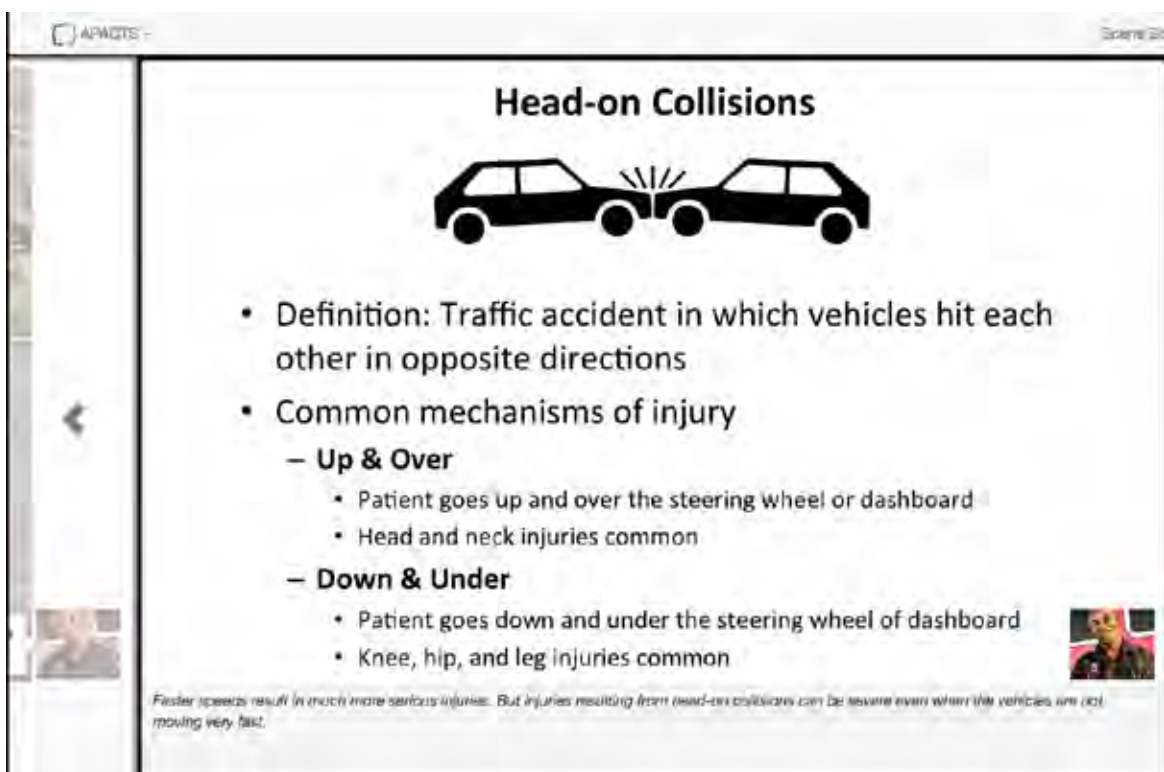
Examples of more detailed/technical knowledge introduced in the study.

Side Impact Collisions

- Definition: Traffic accident in which one vehicle is struck on its side by another

- Common mechanisms of injury
  - Side impact: body often thrown sideways
  - Possible direct, blunt injury on the impacted side
  - Head and neck injuries common

What is the most likely MOI for the driver of the left car?

Another "check your knowledge" question.

Blunt injury from door (due to side impact)

Neck injury

Head-on collision (up and over)

Head-on collision (down and under)

What is the most likely MOI for the driver of the left car?

You are the first and only emergency medical personnel on-site. What is your first action following scene size-up?

This is an example of a more challenging
question for a similar instructional context.

Attend to blue car patients (no air bags).

Talk to police to evaluate speed of collision

Call for backup/additional medical support

Verbally check in with the patients in the right car

Definite head injuries. Scramble helo.

You are the first and only emergency medical personnel on-site. What is your first action following scene size-up?
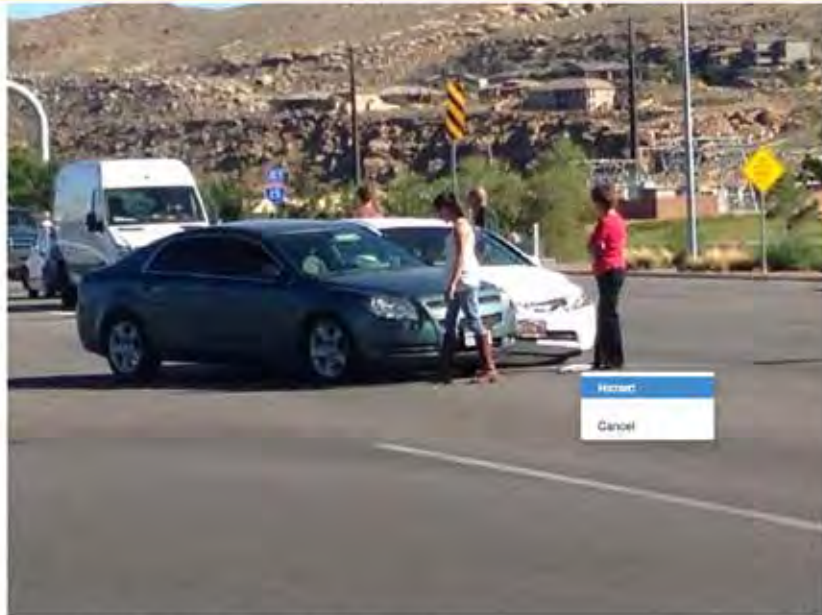
Identify all hazards in the scene.

Adaptation can also include the choice of images with more/less challenging perceptual content. The dog (potential hazard) is easier to perceive in this image than the following one.

Identify all hazards in the scene.

Identify all visible hazards in the image. Press the arrow on the right to move on.

Subjects can also be asked to identify specific areas on an image corresponding to an instructional concept (in this case, identifying hazards).

Identify all visible hazards in the image. Press the arrow on the right to move on.

Appendix E

Recruitment Flyer

Have you ever wondered
what it takes to become an
Emergency Medical Technician?

Participate in this (roughly) 1 hour study to
experience some of the skills EMT's learn!

You will learn a little about being an EMT,
then apply what you've learned in an
interactive learning environment.

You will be compensated $15 for your
participation!

Please contact Kimberly Stowers at
xxxxxxxxx@xxxxxx.xxx to participate.